

# PATH COUPLING AND AGGREGATE PATH COUPLING

YEVGENIY KOVCHEGOV AND PETER T. OTTO

**ABSTRACT.** In this survey paper, we describe and characterize an extension to the classical path coupling method applied statistical mechanical models, referred to as aggregate path coupling. In conjunction with large deviations estimates, we use this aggregate path coupling method to prove rapid mixing of Glauber dynamics for a large class of statistical mechanical models, including models that exhibit discontinuous phase transitions which have traditionally been more difficult to analyze rigorously. The parameter region for rapid mixing for the generalized Curie-Weiss-Potts model is derived as a new application of the aggregate path coupling method.

## CONTENTS

1. Introduction	2
2. Mixing Times and Path Coupling	2
2.1. Coupling Method	3
2.2. Path Coupling	4
2.3. Random-to-Random Shuffling	6
3. Gibbs Ensembles and Glauber Dynamics	9
4. Large Deviations and Equilibrium Macrostate Phase Transitions	10
5. Mean-field Blume-Capel model	12
5.1. Equilibrium Phase Structure	12
5.2. Glauber Dynamics	16
5.3. Path Coupling	17
5.4. Standard Path Coupling in the Continuous Phase Transition Region	21
5.5. Aggregate Path Coupling in the Discontinuous Phase Transition Region	22
5.6. Slow Mixing	24
6. Aggregate Path Coupling for General Class of Gibbs Ensembles	25
6.1. Class of Gibbs Ensembles	26
6.2. Large Deviations	27
6.3. Glauber Dynamics	29
6.4. Coupling of Glauber Dynamics	31
6.5. Mean Coupling Distance	32
6.6. Aggregate Path Coupling	33
6.7. Main Result	36
7. Aggregate Path Coupling applied to the Generalized Potts Model	40
References	47

---

*Date:* January 14, 2015.

2000 *Mathematics Subject Classification.* Primary 60J10; Secondary 60K35.

This work was partially supported by a grant from the Simons Foundation (#284262 to Yevgeniy Kovchegov).

## 1. INTRODUCTION

The theory of mixing times addresses a fundamental question that lies at the heart of statistical mechanics. How quickly does a physical system relax to equilibrium? A related problem arises in computational statistical physics concerning the accuracy of computer simulations of equilibrium data. One typically carries out such simulations by running Glauber dynamics or the closely related Metropolis algorithm, in which case the theory of mixing times allows one to quantify the running time required by the simulation.

An important question driving the work in the field is the relationship between the mixing times of the dynamics and the equilibrium phase transition structure of the corresponding statistical mechanical models. Many results for models that exhibit a continuous phase transition were obtained by a direct application of the standard path coupling method. Standard path coupling [5] is a powerful tool in the theory of mixing times of Markov chains in which rapid mixing can be proved by showing that the mean coupling distance contracts between all neighboring configurations of a minimal path connecting two arbitrary configurations.

For models that exhibit a discontinuous phase transition, the standard path coupling method fails. In this survey paper, we show how to combine aggregate path coupling and large deviation theory to determine the mixing times of a large class of statistical mechanical models, including those that exhibit a discontinuous phase transition. The aggregate path coupling method extends the use of the path coupling technique in the absence of contraction of the mean coupling distance between all neighboring configurations of a statistical mechanical model. The primary objective of this survey is to characterize the assumptions required to apply this new method of aggregate path coupling.

The manuscript is organized as follows: in Section 2, we give a brief overview of mixing times, coupling and path coupling methods, illustrated with a new example of path coupling. Then, beginning in Section 3, we introduce the class of statistical mechanical models considered in the survey. In Sections 5 and 6, we develop and characterize the theory of aggregate path coupling and apply it in Section 7, where we derive the parameter region for rapid mixing for the generalized Curie-Weiss-Potts model that was introduced recently in [25].

## 2. MIXING TIMES AND PATH COUPLING

The mixing time is a measure of the convergence rate of a Markov chain to its stationary distribution and is defined in terms of the *total variation distance* between two distributions  $\mu$  and  $\nu$  defined by

$$\|\mu - \nu\|_{\text{TV}} = \sup_{A \subset \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|$$

Given the convergence of the Markov chain, we define the *maximal distance to stationary* to be

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\text{TV}}$$

where  $P^t(x, \cdot)$  is the transition probability of the Markov chain starting in configuration  $x$  and  $\pi$  is its stationary distribution. Rather than obtaining bounds on  $d(t)$ , it is sometimes easier to

bound the standardized maximal distance defined by

$$(1) \quad \bar{d}(t) := \max_{x, y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}}$$

which satisfies the following result.

**Lemma 2.1.** ([29] Lemma 4.11) *With  $d(t)$  and  $\bar{d}(t)$  defined above, we have*

$$d(t) \leq \bar{d}(t) \leq 2d(t).$$

Given  $\varepsilon > 0$ , the *mixing time* of the Markov chain is defined by

$$t_{\text{mix}}(\varepsilon) = \min\{t : d(t) \leq \varepsilon\}$$

In the modern theory of Markov chains, the interest is in the mixing time as a function of the system size  $n$  and thus, for emphasis, we will often use the notation  $t_{\text{mix}}^{(n)}(\varepsilon)$ . With only a handful of general techniques, rigorous analysis of mixing times is difficult and the proof of exact mixing time asymptotics (with respect to  $n$ ) of even some basic chains remain elusive. See [29] for a survey on the theory of mixing times.

Rates of mixing times are generally categorized into two groups: *rapid mixing* which implies that the mixing time exhibits polynomial growth with respect to the system size, and *slow mixing* which implies that the mixing time grows exponentially with the system size. Determining where a model undergoes rapid mixing is of major importance, as it is in this region that the application of the dynamics is physically feasible.

**2.1. Coupling Method.** The application of coupling (and path coupling) to mixing times of Markov chains begins with the following lemma:

**Lemma 2.2.** *Let  $\mu$  and  $\nu$  be two probability distributions on  $\Omega$ . Then*

$$\|\mu - \nu\|_{\text{TV}} = \inf\{P\{X \neq Y\} : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}$$

This lemma implies that the total variation distance to stationarity, and thus the mixing time, of a Markov chain can be bounded above by the probability  $P(X_t \neq Y_t)$  for a coupling of the Markov chain  $(X_t, Y_t)$  starting in different configurations; i.e.  $(X_0, Y_0) = (\sigma, \tau)$ , or if one of the coupled chains is distributed by the stationary distribution  $\pi$  for all  $t$ .

We run the coupling of the Markov chain, not necessarily independently, until they meet at time  $\tau_c$ . This is called the *coupling time*. After  $\tau_c$ , we run the chains together. We see that  $X_t$  must have the stationary distribution for  $t \geq \tau_c$ , since  $X_t = Y_t$  after coupling.

**Theorem 2.3** (The Coupling Inequality). *Let  $(X_t, Y_t)$  be a coupling of a Markov chain where  $Y_t$  is distributed by the stationary distribution  $\pi$ . The **coupling time** of the Markov chain is defined by*

$$\tau_c := \min\{t : X_t = Y_t\}.$$

*Then, for all initial states  $x$ ,*

$$\|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq P(X_t \neq Y_t) = P[\tau_c > t]$$

*and thus  $\tau_{\text{mix}}(\varepsilon) \leq \mathbb{E}[\tau_c/\varepsilon]$ .*

From the Coupling Inequality, it is clear that in order to use the coupling method to bound the mixing time of a Markov chain, one needs to bound the coupling time for a coupling of the Markov chain starting in *all* pairs of initial states. The advantage of the *path coupling method* described in the next section is that it only requires a bound on couplings starting in certain pairs of initial states.

**2.2. Path Coupling.** The idea of the path coupling method is to view a coupling that starts in configurations  $\sigma$  and  $\tau$  as a sequence of couplings that start in *neighboring* configurations  $(x_i, x_{i+1})$  such that  $(\sigma = x_0, x_1, x_2, \dots, x_r = \tau)$ . Then the contraction of the original coupling distance can be obtained by proving contraction between neighboring configurations which is often easier to show.

Let  $\Omega$  be a finite sample space, and suppose  $(X_t, Y_t)$  is a coupling of a Markov chain on  $\Omega$ . Suppose also there is a neighborhood structure on  $\Omega$ , and suppose it is transitive in the following sense: for any  $x$  and  $y$ , there is a neighbor-to-neighbor path

$$x \sim x_1 \sim x_2 \sim \dots \sim x_{r-1} \sim y,$$

where  $u \sim v$  denotes that sites  $u$  and  $v$  are neighbors.

Let  $d(x, y)$  be a metric over  $\Omega$  such that  $d(x, y) \geq 1$  for any  $x \neq y$ , and

$$d(x, y) = \min_{\rho: x \rightarrow y} \sum_{i=1}^r d(x_{i-1}, x_i),$$

where the minimum is taken over all neighbor-to-neighbor paths

$$\rho: x_0 = x \sim x_1 \sim x_2 \sim \dots \sim x_{r-1} \sim x_r = y$$

of any number of steps  $r$ . Such metric is called *path metric*. Next, we define the *diameter* of the sample space:

$$\text{diam}(\Omega) = \max_{x, y \in \Omega} d(x, y).$$

Finally, the coupling construction allows us to define the *transportation metric* of Kantorovich (1942) as follows:

$$d_K(x, y) := E[d(X_{t+1}, Y_{t+1}) \mid X_t = x, Y_t = y].$$

One can check  $d_K(x, y)$  is a metric over  $\Omega$ .

Path coupling, invented by Bubley and Dyer in 1997, is a method that employs an existing coupling construction in order to bound the mixing time from above. This method in its standard form usually requires certain metric contraction between neighbor sites. Specifically, we require that for any  $x \sim y$ ,

$$(2) \quad d_K(x, y) = E[d(X_{t+1}, Y_{t+1}) \mid X_t = x, Y_t = y] \leq (1 - \delta(\Omega))d(x, y),$$

where  $0 < \delta(\Omega) < 1$  does not depend on  $x$  and  $y$ .

The above contraction inequality (2) has the following implication.

**Proposition 2.4.** *Suppose inequality (2) is satisfied. Then*

$$t_{\text{mix}}(\epsilon) \leq \left\lceil \frac{\log \text{diam}(\Omega) - \log \epsilon}{\delta(\Omega)} \right\rceil.$$

*Proof.* For any  $x$  and  $y$  in  $\Omega$ , consider the path metric minimizing path

$$\rho : x_0 = x \sim x_1 \sim x_2 \sim \dots \sim x_{r-1} \sim x_r = y$$

such that

$$d(x, y) = \sum_{i=1}^r d(x_{i-1}, x_i).$$

Then

$$\begin{aligned} E[d(X_{t+1}, Y_{t+1}) | X_t = x, Y_t = y] &= d_K(x, y) \\ &\leq \sum_{i=1}^r d_K(x_{i-1}, x_i) \\ &\leq (1 - \delta(\Omega)) \sum_{i=1}^r d(x_{i-1}, x_i) \\ &= (1 - \delta(\Omega)) d(x, y). \end{aligned}$$

Hence, after  $t$  iterations,

$$E[d(X_t, Y_t)] \leq (1 - \delta(\Omega))^t d(X_0, Y_0) \leq (1 - \delta(\Omega))^t \text{diam}(\Omega)$$

for any initial  $(X_0, Y_0)$ , and

$$P(X_t \neq Y_t) = P(d(X_t, Y_t) \geq 1) \leq E[d(X_t, Y_t)] \leq (1 - \delta(\Omega))^t \text{diam}(\Omega) \leq \epsilon$$

whenever

$$t \geq \frac{\log \text{diam}(\Omega) - \log \epsilon}{-\log(1 - \delta(\Omega))}.$$

Thus, by the Coupling Inequality,

$$t_{\text{mix}}(\epsilon) \leq \left\lceil \frac{\log \text{diam}(\Omega) - \log \epsilon}{-\log(1 - \delta(\Omega))} \right\rceil \leq \left\lceil \frac{\log \text{diam}(\Omega) - \log \epsilon}{\delta(\Omega)} \right\rceil.$$

□

**Example.** Consider the Ising model over a  $d$ -dimensional torus  $\mathbb{Z}^d/n\mathbb{Z}^d$ . There  $\Omega = \{-1, +1\}^{n^d}$  is the space of all spin configurations, and for any pair of configurations  $\sigma, \tau \in \Omega$ , the path metric  $d(\sigma, \tau)$  is the number of discrepancies between them

$$d(\sigma, \tau) = \sum_{x \in \mathbb{Z}^d/n\mathbb{Z}^d} \mathbf{1}\{\sigma_x \neq \tau_x\}.$$

The diameter  $\text{diam}(\Omega) = n^d$ . It can be checked that if the inverse temperature parameter  $\beta$  satisfies  $\tanh(\beta) < \frac{1}{2d}$ , the contraction inequality (2) is satisfied with

$$\delta(n) = \frac{1 - 2d \tanh(\beta)}{n}.$$

Hence

$$t_{mix}(\epsilon) \leq \left\lceil \frac{\log \text{diam}(\Omega) - \log \epsilon}{\delta(\Omega)} \right\rceil = \left\lceil n \frac{d \log n - \log \epsilon}{1 - 2d \tanh(\beta)} \right\rceil = O(Cn \log n),$$

where  $C = \frac{d}{1 - 2d \tanh(\beta)}$ .

The emergence of the path coupling technique [5] has allowed for a greater simplification in the use of the coupling argument, as rigorous analysis of coupling can be significantly easier when one considers only neighboring configurations. However, the simplification of the path coupling technique comes at the cost of the strong assumption that the coupling distance for all pairs of neighboring configurations must be contracting. Observe that although the contraction between all neighbors is a sufficient condition for the above mixing time bound, it is far from being a necessary condition. In fact, this condition is an artifact of the method.

There had been some successful generalizations of the path coupling method. Specifically in [15], [24] and [4]. In [15], the path coupling method is generalized to account for contraction after a specific number of time-steps, defined as a random variable. In [24] a multi-step non-Markovian coupling construction is considered that evolves via partial couplings of variable lengths determined by stopping times. In order to bound the coupling time, the authors of [24] introduce a technique they call *variable length path coupling* that further generalizes the approach in [15].

**2.3. Random-to-Random Shuffling.** An example illustrating the idea of path coupling can be found in the REU project [32] of Jennifer Thompson that was supervised by Yevgeniy Kovchegov in the summer of 2010 at Oregon State University. There, we consider the shuffling algorithm whereby on each iteration we select a card uniformly from the deck, remove it from the deck, and place it in one of the  $n$  positions in the deck, selected uniformly and independently. Each iteration being done independently of the others. This is referred to as the random-to-random card shuffling algorithm. We need to shuffle the deck so that when we are done with shuffling the deck each of  $n!$  possible permutations is obtained with probability close to  $\frac{1}{n!}$ . Its mixing time can be easily shown to be of order  $O(n \log n)$  using the notion of *strong stationary time*. For this one would consider the time it takes for each card in the deck to be selected at least once. Then use the *coupon collector problem* to prove the  $O(n \log n)$  upper bound on the mixing time. The same coupon collector problem is applied to show that we need at least  $O(n \log n)$  iterations of the shuffling algorithm to mix the deck. The goal of the REU project in [32] was to arrive with the  $O(n \log n)$  upper bound using the coupling method.

**2.3.1. The Coupling.** Take two decks of  $n$  cards, A and B.

- Randomly choose  $i \in [1, n]$ .
- Remove card with label  $i$  from each deck.
- Randomly reinsert card  $i$  in deck A.
- (1) If the new location of  $i$  in A is the top of A, then insert  $i$  on the top of B.
- (2) If the new location of  $i$  in A is below card  $j$ , insert  $i$  below  $j$  in B.

Let  $A_t \in S_n$  and  $B_t \in S_n$  denote the card orderings (permutations) in decks A and B after  $t$  iterations.

<b>A</b>	<b>B</b>
2	4
4	3
1	1
3	2

FIGURE 1. One configuration of matchings between two decks of  $n = 4$  cards.

2.3.2. *Computing the coupling time with a laces approach.* We introduce the following path metric  $d(\cdot, \cdot) : S_n \times S_n \rightarrow \mathbb{Z}_+$  by letting  $d(\sigma, \sigma')$  be the minimal number of nearest neighbour transpositions to traverse between the two permutations,  $\sigma$  and  $\sigma'$ . For example, for the two decks A and B in Figure 1, a distance minimizing path connecting the two permutations is given in Figure 2.

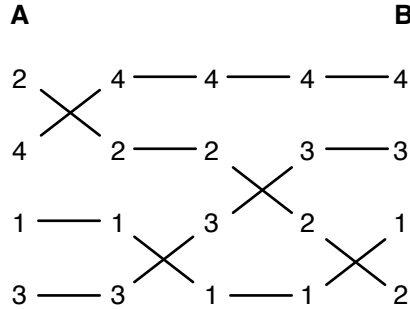


FIGURE 2. Minimal number of crossings between the two permutations is four.

Note that  $d(\sigma, \sigma') \leq \binom{n}{2}$ . We consider the quantity  $d_t = d(A_t, B_t)$ , the distance between our two decks at time  $t$ . We want to find the relationship between  $\mathbb{E}[d_{t+1}]$  and  $\mathbb{E}[d_t]$ .

We consider a  $d(\cdot, \cdot)$ -metric minimizing path. We call the path taken by a card label a *lace*. Thus each lace representing a card label is involved in a certain number of crossings. Let  $r_t$  be the number crossings per lace, averaged over all  $n$  card labels. Then we have  $d_t = \frac{nr_t}{2}$ .

The evolution of the path connecting  $A_t$  to  $B_t$  can be described as following. At each timestep we pick a lace (corresponding to a card label, say  $i$ ) at random and remove it. For example, take a minimal path connecting decks A and B in Figure 2, and remove a lace corresponding to label 3, obtaining Figure 3. Then we reinsert the removed lace back. There will be two cases:

- (1) With probability  $\frac{1}{n}$  we place the lace corresponding to card label  $i$  to the top of the deck. See Figure 4. Then there will be no new crossings.

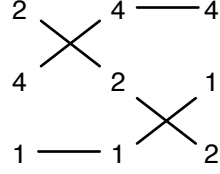


FIGURE 3. Removing lace 3 decreases the number of crossings to two.

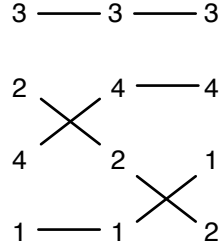


FIGURE 4. Placing lace 3 on top does not add new crossings.

- (2) We choose a lace  $j$  randomly and uniformly chosen among the remaining  $n - 1$  laces, and place lace  $i$  directly below lace  $j$ . This has probability  $\frac{n-1}{n}$ . Then the number of additional new crossings is the same as the number of crossings of lace  $j$ , as in Figure 5. Here

$$\mathbb{E}[\text{new crossings}] = \mathbb{E}[\text{average number of crossings for the remaining laces}] = \left( \frac{nr_t}{2} - r_t \right) \frac{1}{n-1}.$$

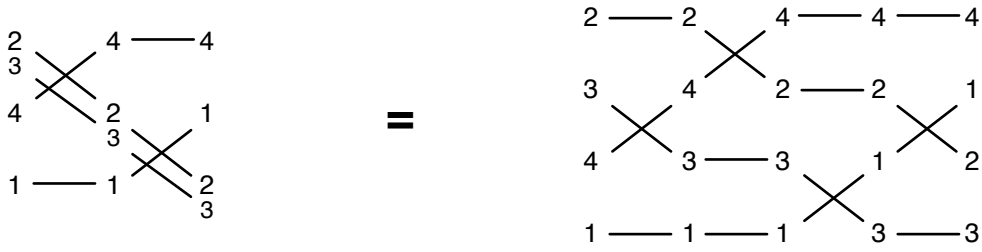


FIGURE 5. Inserting lace 3 directly below lace 2 adds the same number of crossing as there were of lace 2.

Then

$$\mathbb{E}[d_{t+1}|A_t, B_t] = \frac{nr_t}{2} - r_t + \left( \frac{n-1}{n} \right) \left( \frac{nr_t}{2} - r_t \right) \frac{1}{n-1} = \left( 1 - \frac{1}{n} - \frac{2}{n^2} \right) d_t.$$

Hence

$$\mathbb{E}[d_{t+1}] = \left( 1 - \frac{1}{n} - \frac{2}{n^2} \right) \mathbb{E}[d_t],$$



and therefore

$$P(A_t \neq B_t) = P(d_t \geq 1) \leq \mathbb{E}[d_t] = \left(1 - \frac{1}{n} - \frac{2}{n^2}\right)^t \mathbb{E}[d_0] \leq \left(1 - \frac{1}{n} - \frac{2}{n^2}\right)^t \binom{n}{2} \leq \epsilon$$

whenever

$$t \geq \frac{-2 \log n + \log 2 + \log \epsilon}{\log \left(1 - \frac{1}{n} - \frac{2}{n^2}\right)} = 2n \log n + O(n).$$

Thus providing an upper bound on mixing time.

### 3. GIBBS ENSEMBLES AND GLAUBER DYNAMICS

In recent years, mixing times of dynamics of statistical mechanical models have been the focus of much probability research, drawing interest from researchers in mathematics, physics and computer science. The topic is both physically relevant and mathematically rich. But up to now, most of the attention has focused on particular models including rigorous results for several mean-field models. A few examples are (a) the Curie-Weiss (mean-field Ising) model [13, 14, 20, 28], (b) the mean-field Blume-Capel model [19, 26], (c) the Curie-Weiss-Potts (mean-field Potts) model [1, 11]. A good survey of the topic of mixing times of statistical mechanical models can be found in the recent paper by Cuff et. al. [11].

The aggregate path coupling method was developed in [26] and [27] to obtain rapid mixing results for statistical mechanical models, in particular, those models that undergo a first-order, discontinuous phase transition. For this class of models, the standard path coupling method fails to be applicable. The remainder of this survey is devoted to the exposition of the aggregate path coupling method applied to statistical mechanical models.

As stated in [17], “In statistical mechanics, one derives macroscopic properties of a substance from a probability distribution that describes the complicated interactions among the individual constituent particles.” The distribution referred to in this quote is called the Gibbs ensemble or Gibbs measure which are defined next.

A *configuration* of the model has the form  $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Lambda^n$ , where  $\Lambda$  is some finite, discrete set. We will consider a configuration on a graph with  $n$  vertices and let  $X_i(\omega) = \omega_i$  denote the *spin* at vertex  $i$ . The random variables  $X_i$ ’s for  $i = 1, 2, \dots, n$  are independent and identically distributed with common distribution  $\rho$ . The interactions among the spins are defined through the *Hamiltonian* function  $H_n$  and we denote by  $M_n(\omega)$  the relevant macroscopic quantity corresponding to the configuration  $\omega$ . The lift from the microscopic level of the configurations to the macroscopic level of  $M_n$  is through the *interaction representation function*  $H$  that satisfies

$$(3) \quad H_n(\omega) = nH(M_n(\omega)).$$

**Definition 3.1.** *The Gibbs measure or Gibbs ensemble in statistical mechanics is defined as*

$$(4) \quad P_{n,\beta}(B) = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta H_n(\omega)\} dP_n = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta n H(M_n(\omega))\} dP_n$$

where  $P_n$  is the product measure with identical marginals  $\rho$  and  $Z_n(\beta) = \int_{\Lambda^n} \exp \{-\beta H_n(\omega)\} dP_n$  is the **partition function**. The positive parameter  $\beta$  represents the inverse temperature of the external heat bath.

**Definition 3.2.** On the configuration space  $\Lambda^n$ , we define the **Glauber dynamics** for the class of spin models considered in this paper. These dynamics yield a reversible Markov chain  $X^t$  with stationary distribution being the Gibbs ensemble  $P_{n,\beta}$ .

- (i) Select a vertex  $i$  from the underlying graph uniformly,
- (ii) Update the spin at vertex  $i$  according to the distribution  $P_{n,\beta}$ , conditioned on the event that the spins at all vertices not equal to  $i$  remain unchanged.

For more on Glauber dynamics, see [6].

An important question of mixing times of dynamics of statistical mechanical models is its relationship with the thermodynamic phase transition structure of the system. More specifically, as a system undergoes an equilibrium phase transition with respect to some parameter; e.g. temperature, how do the corresponding mixing times behave? The answer to this question depends on the type of thermodynamic phase transition exhibited by the model. In the next section we define the two types of thermodynamic phase transition via the large deviation principle of the macroscopic quantity.

#### 4. LARGE DEVIATIONS AND EQUILIBRIUM MACROSTATE PHASE TRANSITIONS

The application of the aggregate path coupling method to prove rapid mixing takes advantage of large deviations estimates that these models satisfy. In this section, we give a brief summary of large deviations theory used in this paper, written in the context of Gibbs ensembles defined in the previous section. For a more complete theory of large deviations see for example [12] and [17].

A function  $I$  on  $\mathbb{R}^q$  is called a **rate function** if  $I$  maps  $\mathbb{R}^q$  to  $[0, \infty]$  and has compact level sets.

**Definition 4.1.** Let  $I_\beta$  be a rate function on  $\mathbb{R}^q$ . The sequence  $\{M_n\}$  with respect to the Gibbs ensemble  $P_{n,\beta}$  is said to satisfy the **large deviation principle** (LDP) on  $\mathbb{R}^q$  with rate function  $I_\beta$  if the following two conditions hold.

For any closed subset  $F$ ,

$$(5) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_{n,\beta}\{M_n \in F\} \leq -I_\beta(F)$$

and for any open subset  $G$ ,

$$(6) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{n,\beta}\{M_n \in G\} \geq -I_\beta(G)$$

where  $I_\beta(A) = \inf_{z \in A} I_\beta(z)$ .

The LDP upper bound in the above definition implies that values  $z$  satisfying  $I_\beta(z) > 0$  have an exponentially small probability of being observed as  $n \rightarrow \infty$ . Hence we define the set of **equilibrium macrostates** of the system by

$$\mathcal{E}_\beta = \{z : I_\beta(z) = 0\}.$$

For the class of Gibbs ensembles studied in this survey paper, the set of equilibrium macrostates exhibits the following general behavior. There exists a phase transition critical value of the parameter  $\beta_c$  such that

- (a) for  $0 < \beta < \beta_c$ , the set  $\mathcal{E}_\beta$  consists of a single equilibrium macrostate (single phase); i.e.

$$\mathcal{E}_\beta = \{\tilde{z}_\beta\}$$

- (b) for  $\beta > \beta_c$ , the set  $\mathcal{E}_\beta$  consists of a multiple equilibrium macrostates (multiple phase); i.e.

$$\mathcal{E}_\beta = \{z_{\beta,1}, z_{\beta,2}, \dots, z_{\beta,q}\}$$

The transition from the single phase to the multiple phase follows one of two general types.

Continuous, second-order phase transition: For all  $j = 1, 2, \dots, q$ ,  $\lim_{\beta \rightarrow \beta_c^+} z_{\beta,j} = \tilde{z}_\beta$

Discontinuous, first-order phase transition: For all  $j = 1, 2, \dots, q$ ,  $\lim_{\beta \rightarrow \beta_c^+} z_{\beta,j} \neq \tilde{z}_\beta$

As mentioned in the Introduction, understanding the relationship between the mixing times of the Glauber dynamics and the equilibrium phase transition structure of the corresponding Gibbs ensembles is a major motivation for the work discussed in this paper.

Recent rigorous results for statistical mechanical models that undergo continuous, second-order phase transitions, like the famous Ising model, have been published in [29, 28, 13]. For these models, it has been shown that the mixing times undergo a transition at precisely the thermodynamic phase transition point. In order to show rapid mixing in the subcritical parameter regime ( $\beta < \beta_c$ ) for these models, the classical path coupling method can be applied directly.

However, for models that exhibit the other type of phase transition: discontinuous, first-order; e.g. Potts model with  $q > 2$  [33, 10] and the Blume-Capel model [2, 3, 7, 8, 9, 23] with weak interaction, the mixing time transition does not coincide with the thermodynamic equilibrium phase transition.

Discontinuous, first-order phase transitions are more intricate than their counterpart, which makes rigorous analysis of these models traditionally more difficult. Furthermore, the more complex phase transition structure causes certain parameter regimes of the models to fall outside the scope of standard mixing time techniques including the classical path coupling method discussed in subsection 2.2. This was the motivation for the development of the aggregate path coupling method.

In the following two sections, we define and characterize the aggregate path coupling method to two distinct classes of statistical mechanical spin models. We begin with the mean-field Blume-Capel (BC) model, a model ideally suited for the analysis of the relationship between the thermodynamic equilibrium behavior and mixing times due to its intricate phase transition structure. Specifically, the phase diagram of the BC model includes a curve at which the model undergoes a second-order, continuous phase transition, a curve where the model undergoes a first-order, discontinuous phase transition, and a tricritical point which separates the two curves. Moreover, the BC model clearly illustrates the strength of the aggregate path coupling method within the simpler setting where the macroscopic quantity for the model is one dimensional.

In section 6, we generalize the ideas applied to the BC model and define the aggregate path coupling method to a large class of statistical mechanical models with macroscopic quantities

in arbitrary dimensions. We end the survey paper with new mixing time results for Glauber dynamics that converge to the so-called generalized Potts model on the complete graph [25] by applying the general aggregate path coupling method derived in that section .

## 5. MEAN-FIELD BLUME-CAPEL MODEL

The Hamiltonian function on the configuration space  $\Lambda^n = \{-1, 0, 1\}^n$  for the mean-field Blume-Capel model is defined by

$$H_{n,K}(\omega) = \sum_{j=1}^n \omega_j^2 - \frac{K}{n} \left( \sum_{j=1}^n \omega_j \right)^2$$

for configurations  $\omega = (\omega_1, \dots, \omega_n)$ . Here  $K$  represents the interaction strength of the model. Then for inverse temperature  $\beta$ , the mean-field Blume-Capel model is defined by the sequence of probability measures

$$P_{n,\beta,K}(\omega) = \frac{1}{Z_n(\beta, K)} \exp[-\beta H_{n,K}(\omega)]$$

where  $Z_n(\beta, K) = \sum_{\omega \in \Lambda^n} \exp[-\beta H_{n,K}(\omega)]$  is the normalizing constant called the *partition function*.

**5.1. Equilibrium Phase Structure.** In [23], using large deviation theory [18], the authors proved the phase transition structure of the BC model. The analysis of  $P_{n,\beta,K}$  was facilitated by expressing it in the form of a Curie-Weiss (mean-field Ising)-type model. This is done by absorbing the noninteracting component of the Hamiltonian into the product measure  $P_n$  that assigns the probability  $3^{-n}$  to each  $\omega \in \Lambda^n$ , obtaining

$$(7) \quad P_{n,\beta,K}(d\omega) = \frac{1}{\tilde{Z}_n(\beta, K)} \cdot \exp \left[ n\beta K \left( \frac{S_n(\omega)}{n} \right)^2 \right] P_{n,\beta}(d\omega)$$

In this formula  $S_n(\omega)$  equals the total spin  $\sum_{j=1}^n \omega_j$ ,  $P_{n,\beta}$  is the product measure on  $\Lambda^n$  with identical one-dimensional marginals

$$(8) \quad \rho_\beta(d\omega_j) = \frac{1}{Z(\beta)} \cdot \exp(-\beta \omega_j^2) \rho(d\omega_j),$$

$Z(\beta)$  is the normalizing constant  $\int_{\Lambda} \exp(-\beta \omega_j^2) \rho(d\omega_j) = 1 + 2e^{-\beta}$ , and  $\tilde{Z}_n(\beta, K)$  is the normalizing constant  $[Z(\beta)]^n / Z_n(\beta, K)$ .

Although  $P_{n,\beta,K}$  has the form of a Curie-Weiss (mean-field Ising) model when rewritten as in (7), it is much more complicated because of the  $\beta$ -dependent product measure  $P_{n,\beta}$  and the presence of the parameter  $K$ . These complications introduce new features to the BC model described above that are not present in the Curie-Weiss model [17].

The starting point of the analysis of the phase-transition structure of the BC model is the large deviation principle (LDP) satisfied by the spin per site or *magnetization*  $S_n/n$  with respect to

$P_{n,\beta,K}$ . In order to state the form of the rate function, we introduce the cumulant generating function  $c_\beta$  of the measure  $\rho_\beta$  defined in (8); for  $t \in \mathbb{R}$  this function is defined by

$$c_\beta(t) = \log \int_{\Lambda} \exp(t\omega_1) \rho_\beta(d\omega_1) = \log \left[ \frac{1 + e^{-\beta}(e^t + e^{-t})}{1 + 2e^{-\beta}} \right]$$

We also introduce the Legendre-Fenchel transform of  $c_\beta$ , which is defined for  $z \in [-1, 1]$  by

$$J_\beta(z) = \sup_{t \in \mathbb{R}} \{tz - c_\beta(t)\}$$

and is finite for  $z \in [-1, 1]$ .  $J_\beta$  is the rate function in Cramér's theorem, which is the LDP for  $S_n/n$  with respect to the product measures  $P_{n,\beta}$  [17, Thm. II.4.1] and is one of the components of the proof of the LDP for  $S_n/n$  with respect to the BC model  $P_{n,\beta,K}$ . This LDP is stated in the next theorem and is proved in Theorem 3.3 in [23].

**Theorem 5.1.** *For all  $\beta > 0$  and  $K > 0$ , with respect to  $P_{n,\beta,K}$ ,  $S_n/n$  satisfies the large deviation principle on  $[-1, 1]$  with exponential speed  $n$  and rate function*

$$I_{\beta,K}(z) = J_\beta(z) - \beta K z^2 - \inf_{y \in \mathbb{R}} \{J_\beta(y) - \beta K y^2\}.$$

The LDP in the above theorem implies that those  $z \in [-1, 1]$  satisfying  $I_{\beta,K}(z) > 0$  have an exponentially small probability of being observed as  $n \rightarrow \infty$ . Hence we define the set of equilibrium macrostates by

$$\tilde{\mathcal{E}}_{\beta,K} = \{z \in [-1, 1] : I_{\beta,K}(z) = 0\}.$$

For  $z \in \mathbb{R}$  we define

$$(9) \quad G_{\beta,K}(z) = \beta K z^2 - c_\beta(2\beta K z)$$

and as in [21] and [22] refer to it as the *free energy functional* of the model. The calculation of the zeroes of  $I_{\beta,K}$  — equivalently, the global minimum points of  $J_{\beta,K}(z) - \beta K z^2$  — is greatly facilitated by the following observations made in Proposition 3.4 in [23]:

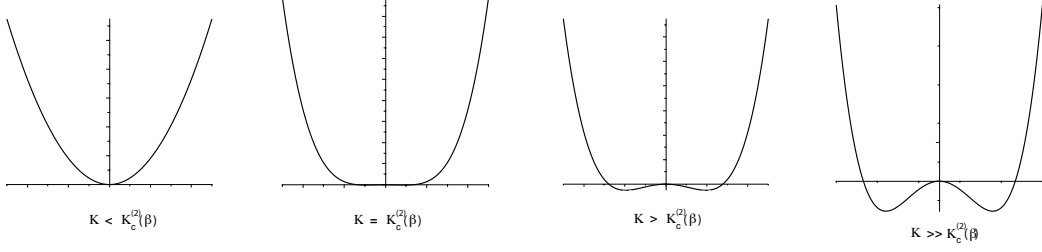
- (1) The global minimum points of  $J_{\beta,K}(z) - \beta K z^2$  coincide with the global minimum points of  $G_{\beta,K}$ , which are much easier to calculate.
- (2) The minimum values  $\min_{z \in \mathbb{R}} \{J_{\beta,K}(z) - \beta K z^2\}$  and  $\min_{z \in \mathbb{R}} \{G_{\beta,K}(z)\}$  coincide.

Item (1) gives the alternate characterization that

$$(10) \quad \tilde{\mathcal{E}}_{\beta,K} = \{z \in [-1, 1] : z \text{ minimizes } G_{\beta,K}(z)\}.$$

The free energy functional  $G_{\beta,K}$  exhibits two distinct behaviors depending on whether  $\beta \leq \beta_c = \log 4$  or  $\beta > \beta_c$ . In the first case, the behavior is similar to the Curie-Weiss (mean-field Ising) model. Specifically, there exists a critical value  $K_c^{(2)}(\beta)$  defined in (11) such that for  $K < K_c^{(2)}(\beta)$ ,  $G_{\beta,K}$  has a single minimum point at  $z = 0$ . At the critical value  $K = K_c^{(2)}(\beta)$ ,  $G_{\beta,K}$  develops symmetric non-zero minimum points and a local maximum point at  $z = 0$ . This behavior corresponds to a continuous, second-order phase transition and is illustrated in Figure 6.

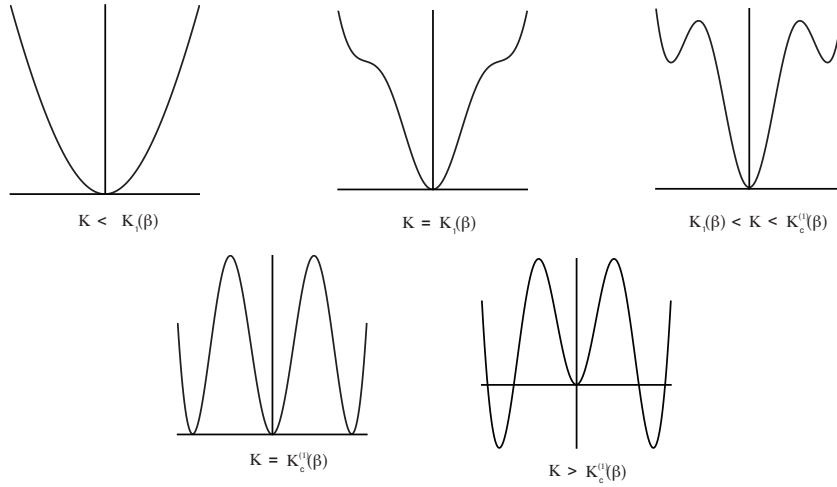
On the other hand, for  $\beta > \beta_c$ ,  $G_{\beta,K}$  undergoes two transitions at the values denoted by  $K_1(\beta)$  and  $K_c^{(1)}(\beta)$ . For  $K < K_1(\beta)$ ,  $G_{\beta,K}$  again possesses a single minimum point at  $z = 0$ . At the

FIGURE 6. The free-energy functional  $G_{\beta,K}$  for  $\beta \leq \beta_c$ 

first critical value  $K_1(\beta)$ ,  $G_{\beta,K}$  develops symmetric non-zero local minimum points in addition to the global minimum point at  $z = 0$ . These local minimum points are referred to as *metastable states* and we refer to  $K_1(\beta)$  as the *metastable critical value*. This value is defined implicitly in Lemma 3.9 of [23] as the unique value of  $K$  for which there exists a unique  $z > 0$  such that

$$G'_{\beta,K_1(\beta)}(z) = 0 \quad \text{and} \quad G''_{\beta,K_1(\beta)}(z) = 0$$

As  $K$  increases from  $K_1(\beta)$  to  $K_c^{(1)}(\beta)$ , the local minimum points decrease until at  $K = K_c^{(1)}(\beta)$ , the local minimum points reach zero and  $G_{\beta,K}$  possesses three global minimum points. Therefore, for  $\beta > \beta_c$ , the BC model undergoes a phase transition at  $K = K_c^{(1)}(\beta)$ , which is defined implicitly in [23]. Lastly, for  $K > K_c^{(1)}(\beta)$ , the symmetric non-zero minimum points drop below zero and thus  $G_{\beta,K}$  has two symmetric non-zero global minimum points. This behavior corresponds to a discontinuous, first-order phase transition and is illustrated in Figure 7.

FIGURE 7. The free-energy functional  $G_{\beta,K}$  for  $\beta > \beta_c$

In the next two theorems, the structure of  $\tilde{\mathcal{E}}_{\beta,K}$  corresponding to the behavior of  $G_{\beta,K}$  just described is stated which depends on the relationship between  $\beta$  and the critical value  $\beta_c = \log 4$ . We first describe  $\tilde{\mathcal{E}}_{\beta,K}$  for  $0 < \beta \leq \beta_c$  and then for  $\beta > \beta_c$ . In the first case  $\tilde{\mathcal{E}}_{\beta,K}$  undergoes a continuous bifurcation as  $K$  increases through the critical value  $K_c^{(2)}(\beta)$  defined in (11); physically, this bifurcation corresponds to a second-order phase transition. The following theorem is proved in Theorem 3.6 in [23].

**Theorem 5.2.** *For  $0 < \beta \leq \beta_c$ , we define*

$$(11) \quad K_c^{(2)}(\beta) = \frac{1}{2\beta c''_\beta(0)} = \frac{e^\beta + 2}{4\beta}.$$

*For these values of  $\beta$ ,  $\tilde{\mathcal{E}}_{\beta,K}$  has the following structure.*

- (a) *For  $0 < K \leq K_c^{(2)}(\beta)$ ,  $\tilde{\mathcal{E}}_{\beta,K} = \{0\}$ .*
- (b) *For  $K > K_c^{(2)}(\beta)$ , there exists  $z(\beta, K) > 0$  such that  $\tilde{\mathcal{E}}_{\beta,K} = \{\pm z(\beta, K)\}$ .*
- (c)  *$z(\beta, K)$  is a positive, increasing, continuous function for  $K > K_c^{(2)}(\beta)$ , and as  $K \rightarrow (K_c^{(2)}(\beta))^+$ ,  $z(\beta, K) \rightarrow 0$ . Therefore,  $\tilde{\mathcal{E}}_{\beta,K}$  exhibits a continuous bifurcation at  $K_c^{(2)}(\beta)$ .*

For  $\beta \in (0, \beta_c)$ , the curve  $(\beta, K_c^{(2)}(\beta))$  is the curve of second-order critical points. As we will see in a moment, for  $\beta \in (\beta_c, \infty)$  the BC model also has a curve of first-order critical points, which we denote by  $(\beta, K_c^{(1)}(\beta))$ .

We now describe  $\tilde{\mathcal{E}}_{\beta,K}$  for  $\beta > \beta_c$ . In this case  $\tilde{\mathcal{E}}_{\beta,K}$  undergoes a discontinuous bifurcation as  $K$  increases through an implicitly defined critical value. Physically, this bifurcation corresponds to a first-order phase transition. The following theorem is proved in Theorem 3.8 in [23].

**Theorem 5.3.** *For all  $\beta > \beta_c$ ,  $\tilde{\mathcal{E}}_{\beta,K}$  has the following structure in terms of the quantity  $K_c^{(1)}(\beta)$  defined implicitly for  $\beta > \beta_c$  on page 2231 of [23].*

- (a) *For  $0 < K < K_c^{(1)}(\beta)$ ,  $\tilde{\mathcal{E}}_{\beta,K} = \{0\}$ .*
- (b) *There exists  $z(\beta, K_c^{(1)}(\beta)) > 0$  such that  $\tilde{\mathcal{E}}_{\beta, K_c^{(1)}(\beta)} = \{0, \pm z(\beta, K_c^{(1)}(\beta))\}$ .*
- (c) *For  $K > K_c^{(1)}(\beta)$  there exists  $z(\beta, K) > 0$  such that  $\tilde{\mathcal{E}}_{\beta,K} = \{\pm z(\beta, K)\}$ .*
- (d)  *$z(\beta, K)$  is a positive, increasing, continuous function for  $K \geq K_c^{(1)}(\beta)$ , and as  $K \rightarrow K_c^{(1)}(\beta)^+$ ,  $z(\beta, K) \rightarrow z(\beta, K_c^{(1)}(\beta)) > 0$ . Therefore,  $\tilde{\mathcal{E}}_{\beta,K}$  exhibits a discontinuous bifurcation at  $K_c^{(1)}(\beta)$ .*

The phase diagram of the BC model is depicted in Figure 8. The LDP stated in Theorem 5.1 implies the following weak convergence result used in the proof of rapid mixing in the first-order, discontinuous phase transition region. It is part (a) of Theorem 6.5 in [23].

**Theorem 5.4.** *For  $\beta$  and  $K$  for which  $\tilde{\mathcal{E}}_{\beta,K} = \{0\}$ ,*

$$P_{n,\beta,K}\{S_n/n \in dx\} \implies d_0 \quad \text{as } n \rightarrow \infty.$$

We end this section with a final result that was not included in the original paper [23] but will be used in the proof of the slow mixing result for the BC model. The result states that not only do the global minimum point of  $G_{\beta,K}$  and  $I_{\beta,K}$  coincide, but so do the local minimum points.

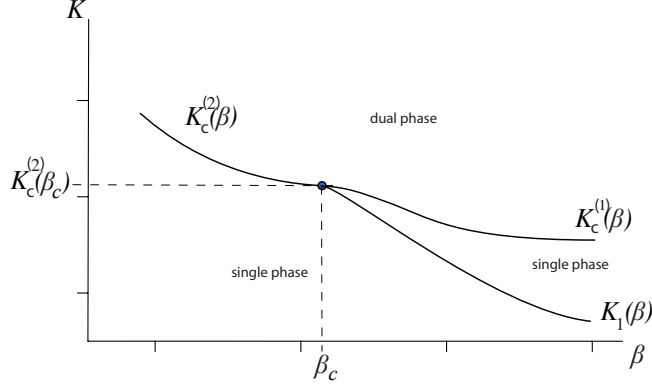


FIGURE 8. Equilibrium phase transition structure of the mean-field Blume-Capel model

**Lemma 5.5.** *In the case where  $G_{\beta,K}$  and  $I_{\beta,K}$  are strictly convex at their minimum points, a point  $\tilde{z}$  is a local minimum point of  $G_{\beta,K}$  if and only if it is a local minimum point of  $I_{\beta,K}$ .*

*Proof.* Assume that  $\tilde{z}$  is a local minimum point of  $G_{\beta,K}$ . Then  $\tilde{z}$  is a critical point of  $G_{\beta,K}$  which implies that  $\tilde{z} = c'_\beta(2\beta K\tilde{z})$ . By the theory of Legendre-Fenchel transforms,  $J'_\beta(z) = (c'_\beta)^{-1}(z)$  and thus

$$I'_{\beta,K}(\tilde{z}) = J'_\beta(\tilde{z}) - 2\beta K\tilde{z} = (c'_\beta)^{-1}(\tilde{z}) - 2\beta K\tilde{z} = 0.$$

Next, since  $\tilde{z}$  is a local minimum point of  $G_{\beta,K}$ ,

$$G''_{\beta,K}(\tilde{z}) > 0 \quad \Longleftrightarrow \quad c''_\beta(2\beta K\tilde{z}) < \frac{1}{2\beta K}$$

Therefore,

$$I''_{\beta,K}(\tilde{z}) = J''_\beta(\tilde{z}) - 2\beta K = \frac{1}{c''_\beta(2\beta K\tilde{z})} - 2\beta K > 0$$

and we conclude that  $\tilde{z}$  is a local minimum point of  $I_{\beta,K}$ . The other direction is obtained by reversing the argument.  $\square$

**5.2. Glauber Dynamics.** The Glauber dynamics, defined in general in section 3, for the mean-field Blume-Capel model evolve by selecting a vertex  $i$  at random and updating the spin at  $i$  according to the distribution  $P_{n,\beta,K}$ , conditioned to agree with the spins at all vertices not equal to  $i$ . If the current configuration is  $\omega$  and vertex  $i$  is selected, then the chance of the spin at  $i$  is updated to  $+1$  is equal to

$$(12) \quad p_{+1}(\omega, i) = \frac{e^{2\beta K \tilde{S}(\omega, i)/n}}{e^{2\beta K \tilde{S}(\omega, i)/n} + e^{\beta - (\beta K)/n} + e^{-2\beta K \tilde{S}(\omega, i)/n}}$$



where  $\tilde{S}(\omega, i) = \sum_{j \neq i} \omega_j$  is the total spin of the neighboring vertices of  $i$ . Similarly, the probabilities of  $i$  updating to 0 and  $-1$  are

$$(13) \quad p_0(\omega, i) = \frac{e^{\beta - (\beta K)/n}}{e^{2\beta K \tilde{S}(\omega, i)/n} + e^{\beta - (\beta K)/n} + e^{-2\beta K \tilde{S}(\omega, i)/n}}$$

and

$$(14) \quad p_{-1}(\omega, i) = \frac{e^{-2\beta K \tilde{S}(\omega, i)/n}}{e^{2\beta K \tilde{S}(\omega, i)/n} + e^{\beta - (\beta K)/n} + e^{-2\beta K \tilde{S}(\omega, i)/n}}$$

$p_{+1}(\omega, i)$  is increasing with respect to  $\tilde{S}(\omega, i)$ ,  $p_{-1}(\omega, i)$  is decreasing with respect to  $\tilde{S}(\omega, i)$ , and  $p_0(\omega, i)$  is decreasing for  $\tilde{S}(\omega, i) > 0$  and increasing for  $\tilde{S}(\omega, i) < 0$ .

A classical tool in proving rapid mixing for Markov chains defined on graphs, including the Glauber dynamics of statistical mechanical models, is the path coupling technique discussed in subsection 2.2. It will be shown that this technique can be directly applied to the BC model in the second-order, continuous phase transition region but fails in a subset of the first-order, discontinuous phase transition region. It is for the latter region that we developed the *aggregate path coupling* method to prove rapid mixing. First, the standard path coupling method for the BC model is introduced in the next section.

**5.3. Path Coupling.** We begin by setting up the coupling rules for the Glauber dynamics of the mean-field Blume-Capel model. Define the path metric  $\rho$  on  $\Omega^n = \{-1, 0, 1\}^n$  by

$$(15) \quad \rho(\sigma, \tau) = \sum_{j=1}^n |\sigma_j - \tau_j|.$$

**Remark 5.6.** *In the original paper [26] on the mixing times of the mean-field Blume-Capel model, the incorrect path metric was used. In that paper, the path metric was defined by*

$$\tilde{\rho}(\sigma, \tau) = \sum_{j=1}^n \mathbf{1}\{\sigma_j \neq \tau_j\}$$

*With the correct metric defined in (15), the proofs in [26] remains the same.*

Let  $\sigma$  and  $\tau$  be two configurations with  $\rho(\sigma, \tau) = 1$ ; i.e.  $\sigma$  and  $\tau$  are neighboring configurations. The spins of  $\sigma$  and  $\tau$  agree everywhere except at a single vertex  $i$ , where either  $\sigma_i = 0$  and  $\tau_i \neq 0$ , or  $\sigma_i \neq 0$  and  $\tau_i = 0$ . Assume that  $\sigma_i = 0$  and  $\tau_i = 1$ . We next describe the path coupling  $(X, Y)$  of one step of the Glauber dynamics starting in configuration  $\sigma$  with one starting in configuration  $\tau$ . Pick a vertex  $k$  uniformly at random. We use a single random variable as the common source of noise to update both chains, so the two chains agree as often as possible. In particular, let  $U$  be a uniform random variable on  $[0, 1]$  and set

$$X(k) = \begin{cases} -1 & \text{if } 0 \leq U \leq p_{-1}(\sigma, k) \\ 0 & \text{if } p_{-1}(\sigma, k) < U \leq p_{-1}(\sigma, k) + p_0(\sigma, k) \\ +1 & \text{if } p_{-1}(\sigma, k) + p_0(\sigma, k) < U \leq 1 \end{cases}$$

and

$$Y(k) = \begin{cases} -1 & \text{if } 0 \leq U \leq p_{-1}(\tau, k) \\ 0 & \text{if } p_{-1}(\tau, k) < U \leq p_{-1}(\tau, k) + p_0(\tau, k) \\ +1 & \text{if } p_{-1}(\tau, k) + p_0(\tau, k) < U \leq 1 \end{cases}$$

Set  $X(j) = \sigma_j$  and  $Y(j) = \tau_j$  for  $j \neq k$ .

Since  $\sigma_i < \tau_i$ , for all  $j \neq i$ ,  $\tilde{S}(\sigma, j) < \tilde{S}(\tau, j)$  and thus

$$p_{+1}(\tau, k) > p_{+1}(\sigma, k) \quad \text{and} \quad p_{-1}(\tau, k) < p_{-1}(\sigma, k)$$

The path metric  $\rho$  on the coupling above takes on the following possible values.

$$\rho(X, Y) = \begin{cases} 0 & \text{if } k = i \\ 1 & \text{if } k \neq i \text{ and both chains updates the same} \\ 2 & \text{if } k \neq i \text{ and the chains update differently} \end{cases}$$

Note that since  $\sigma$  and  $\tau$  are neighbor configurations,  $\rho(X, Y) \neq 3$  because the update probabilities of  $X$  and  $Y$  are sufficiently close.

The application of the path coupling technique to prove rapid mixing is dependent on whether the mean coupling distance with respect to the path metric  $\rho$ , denoted by  $\mathbb{E}_{\sigma, \tau}[\rho(X, Y)]$ , contracts over *all* pairs of neighboring configurations.

In the lemma below and following corollary, we derive a working form for the mean coupling distance.

**Lemma 5.7.** *Let  $\rho$  be the path metric defined in (32) and  $(X, Y)$  be the path coupling of one step of the Glauber dynamics of the mean-field Blume-Capel model where  $X$  and  $Y$  start in neighboring configurations  $\sigma$  and  $\tau$ . Define*

$$(16) \quad \varphi_{\beta, K}(x) = \frac{2 \sinh(\frac{2\beta K}{n}x)}{2 \cosh(\frac{2\beta K}{n}x) + e^{\beta - \frac{\beta K}{n}}}$$

Then

$$\mathbb{E}_{\sigma, \tau}[\rho(X, Y)] = \frac{n-1}{n} + \frac{(n-1)}{n} [\varphi_{\beta, K}(S_n(\tau)) - \varphi_{\beta, K}(S_n(\sigma))] + O\left(\frac{1}{n^2}\right)$$

*Proof.* Let  $n_{-1}, n_0$  and  $n_{+1}$  denote the number of  $-1, 0$  and  $+1$  spins, respectively, in configuration  $\sigma$ , not including the spin at vertex  $i$ , where the configurations differ. Note that  $n_{-1} + n_0 + n_{+1} = n - 1$ .

Define  $\varepsilon(-1)$  to be the probability that  $X$  and  $Y$  update differently when the chosen vertex  $k \neq i$  is a  $-1$  spin. Similarly, define  $\varepsilon(0)$  and  $\varepsilon(+1)$ . Then the mean coupling distance can be expressed as

$$\begin{aligned} \mathbb{E}_{\sigma, \tau}[\rho(X, Y)] &= \frac{n-1}{n}(1 - \varepsilon(-1)) + \frac{n_0}{n}(1 - \varepsilon(0)) + \frac{n_{+1}}{n}(1 - \varepsilon(+1)) \\ &\quad + 2 \left[ \frac{n-1}{n}\varepsilon(-1) + \frac{n_0}{n}\varepsilon(0) + \frac{n_{+1}}{n}\varepsilon(+1) \right] \\ &= \frac{n-1}{n} + \frac{n-1}{n}\varepsilon(-1) + \frac{n_0}{n}\varepsilon(0) + \frac{n_{+1}}{n}\varepsilon(+1) \end{aligned}$$

The probability that  $X$  and  $Y$  update differently when the chosen vertex  $k \neq i$  is a  $-1$  spin is given by

$$\begin{aligned}
\varepsilon(-1) &= [p_{-1}(\sigma, k) - p_{-1}(\tau, k)] + [(p_{-1}(\sigma, k) + p_0(\sigma, k)) - (p_{-1}(\tau, k) + p_0(\tau, k))] \\
&= [p_{+1}(\tau, k) - p_{+1}(\sigma, k)] + [p_{-1}(\sigma, k) - p_{-1}(\tau, k)] \\
&= [p_{+1}(\tau, k) - p_{-1}(\tau, k)] + [p_{-1}(\sigma, k) - p_{+1}(\sigma, k)] \\
&= \frac{2 \sinh(\frac{2\beta K}{n}(S_n(\tau) + 1))}{2 \cosh(\frac{2\beta K}{n}(S_n(\tau) + 1)) + e^{\beta - \frac{\beta K}{n}}} - \frac{2 \sinh(\frac{2\beta K}{n}(S_n(\sigma) + 1))}{2 \cosh(\frac{2\beta K}{n}(S_n(\sigma) + 1)) + e^{\beta - \frac{\beta K}{n}}} \\
&= \varphi_{\beta, K}((S_n(\tau) + 1)) - \varphi_{\beta, K}((S_n(\sigma) + 1)) \\
&= \varphi_{\beta, K}(S_n(\tau)) - \varphi_{\beta, K}(S_n(\sigma)) + O\left(\frac{1}{n^2}\right)
\end{aligned}$$

Similarly, we have

$$\varepsilon(0) = \varphi_{\beta, K}(S_n(\tau)) - \varphi_{\beta, K}(S_n(\sigma))$$

and

$$\varepsilon(+1) = \varphi_{\beta, K}((S_n(\tau) - 1)) - \varphi_{\beta, K}((S_n(\sigma) - 1)) = \varphi_{\beta, K}(S_n(\tau)) - \varphi_{\beta, K}(S_n(\sigma)) + O\left(\frac{1}{n^2}\right)$$

and the proof is complete.  $\square$

For  $c_\beta$  defined in (9), we have

$$\varphi_{\beta, K}(x) = c'_\beta \left( \frac{2\beta K}{n} x \right) (1 + O(1/n))$$

which yields the following corollary.

**Corollary 5.8.** *Let  $\rho$  be the path metric defined in (32) and  $(X, Y)$  be the path coupling where  $X$  and  $Y$  start in neighboring configurations  $\sigma$  and  $\tau$ . Then*

$$\mathbb{E}_{\sigma, \tau}[\rho(X, Y)] = \frac{n-1}{n} + \frac{(n-1)}{n} \left[ c'_\beta \left( 2\beta K \frac{S_n(\tau)}{n} \right) - c'_\beta \left( 2\beta K \frac{S_n(\sigma)}{n} \right) \right] + O\left(\frac{1}{n^2}\right).$$

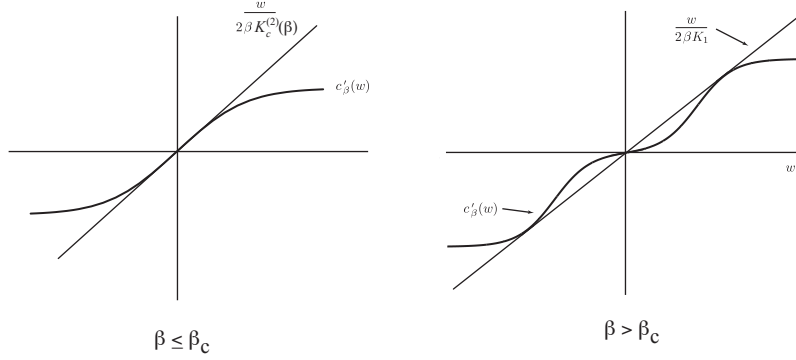
By the above corollary, we conclude that the mean coupling distance of a coupling starting in neighboring configurations contracts; i.e.  $\mathbb{E}_{\sigma, \tau}[\rho(X, Y)] < \rho(\sigma, \tau) = 1$ , if

$$\left[ c'_\beta \left( 2\beta K \frac{S_n(\tau)}{n} \right) - c'_\beta \left( 2\beta K \frac{S_n(\sigma)}{n} \right) \right] \approx 2\beta K \left[ \frac{S_n(\tau)}{n} - \frac{S_n(\sigma)}{n} \right] c''_\beta \left( 2\beta K \frac{S_n(\sigma)}{n} \right) < \frac{1}{n-1}$$

Since  $\sigma$  and  $\tau$  are neighboring configurations and  $S_n(\tau) > S_n(\sigma)$ , this is equivalent to

$$(17) \quad c''_\beta \left( 2\beta K \frac{S_n(\sigma)}{n} \right) < \frac{1}{2\beta K}$$

Therefore, contraction of the mean coupling distance, and thus rapid mixing, depends on the concavity behavior of the function  $c'_\beta$ . This is also precisely what determines the type of thermodynamic equilibrium phase transition (continuous, second-order versus discontinuous, first-order) that is exhibited by the mean-field Blume-Capel model. We state the concavity behavior of  $c'_\beta$  in the next theorem which is proved in Theorem 3.5 in [23]. The results of the theorem are depicted in Figure 9

FIGURE 9. Behavior of  $c'_\beta(w)$  for large and small  $\beta$ .

**Theorem 5.9.** For  $\beta > \beta_c = \log 4$  define

$$(18) \quad w_c(\beta) = \cosh^{-1} \left( \frac{1}{2} e^\beta - 4e^{-\beta} \right) \geq 0.$$

The following conclusions hold.

- (a) For  $0 < \beta \leq \beta_c$ ,  $c'_\beta(w)$  is strictly concave for  $w > 0$ .
- (b) For  $\beta > \beta_c$ ,  $c'_\beta(w)$  is strictly convex for  $0 < w < w_c(\beta)$  and  $c'_\beta(w)$  is strictly concave for  $w > w_c(\beta)$ .

By part (a) of the above theorem, for  $\beta \leq \beta_c$ ,  $c''_\beta(x) \leq c''_\beta(0) = 1/(2\beta K_c^{(2)}(\beta))$ . Therefore, by (17), the mean coupling distance contracts between *all* pairs of neighboring states whenever  $K < K_c^{(2)}(\beta)$ .

By contrast, for  $\beta > \beta_c$ , we will show that rapid mixing occurs whenever  $K < K_1(\beta)$  where  $K_1(\beta)$  is the metastable critical value introduced in Subsection 5.1 and depicted in Figure 7. However, since the supremum  $\sup_{[-1,1]} c''_\beta(x) > \frac{1}{2\beta K_1(\beta)}$ , the condition  $K < K_1(\beta)$  is not sufficient for (17) to hold. That is,  $K < K_1(\beta)$  does not imply the contraction of the mean coupling distance between *all* pairs of neighboring states. However, we prove rapid mixing for all  $K < K_1(\beta)$  in Subsection 5.5 by using an extension to the path coupling method that we refer to as *aggregate path coupling*.

We now prove the mixing times for the mean-field Blume-Capel model, which varies depending on the parameter values  $(\beta, K)$  and their position with respect to the thermodynamic phase transition curves. We begin with the case  $\beta \leq \beta_c$  where the model undergoes a continuous, second-order phase transition and  $K \leq K_c^{(2)}(\beta)$  which corresponds to the single phase region.

**5.4. Standard Path Coupling in the Continuous Phase Transition Region.** We begin by stating the standard path coupling argument used to prove rapid mixing for the mean-field Blume-Capel model in the continuous, second-order phase transition region. The result is proved in Proposition 2.4.

**Theorem 5.10.** *Suppose the state space  $\Omega$  of a Markov chain is the vertex set of a graph with path metric  $\rho$ . Suppose that for each edge  $\{\sigma, \tau\}$  there exists a coupling  $(X, Y)$  of the distributions  $P(\sigma, \cdot)$  and  $P(\tau, \cdot)$  such that*

$$\mathbb{E}_{\sigma, \tau}[\rho(X, Y)] \leq \rho(\sigma, \tau)e^{-\alpha} \quad \text{for some } \alpha > 0$$

Then

$$t_{\text{mix}}(\varepsilon) \leq \left\lceil \frac{-\log(\varepsilon) + \log(\text{diam}(\Omega))}{\alpha} \right\rceil$$

In this section, we assume  $\beta \leq \beta_c$  which implies that the BC model undergoes a continuous, second-order phase transition at  $K = K_c^{(2)}(\beta)$  defined in (11). By Theorem 5.9, for  $\beta \leq \beta_c$ ,  $c'_\beta(x)$  is concave for  $x > 0$ . See the first graph of Figure 9 as reference. We next state and prove the rapid mixing result for the mean-field Blume-Capel model in the second-order, continuous phase transition regime.

**Theorem 5.11.** *Let  $t_{\text{mix}}(\varepsilon)$  be the mixing time for the Glauber dynamics of the mean-field Blume-Capel model on  $n$  vertices and  $K_c^{(2)}(\beta)$  the continuous phase transition curve defined in (11). Then for  $\beta \leq \beta_c = \log 4$  and  $K < K_c^{(2)}(\beta)$ ,*

$$t_{\text{mix}}(\varepsilon) \leq \frac{n}{\alpha}(\log n + \log(1/\varepsilon))$$

for any  $\alpha \in \left(0, \frac{K_c^{(2)}(\beta) - K}{K_c^{(2)}(\beta)}\right)$  and  $n$  sufficiently large.

*Proof.* Let  $(X, Y)$  be a coupling of the Glauber dynamics of the BC model that begin in neighboring configurations  $\sigma$  and  $\tau$  with respect to the path metric  $\rho$  defined in (32). By Corollary 5.8 of Lemma 5.7,

$$\mathbb{E}_{\sigma, \tau}[\rho(X, Y)] = 1 - \left( \frac{1}{n} - \frac{(n-1)}{n} \left[ c'_\beta \left( 2\beta K \frac{S_n(\tau)}{n} \right) - c'_\beta \left( 2\beta K \frac{S_n(\sigma)}{n} \right) \right] \right) + O\left(\frac{1}{n^2}\right)$$

Observe that  $c''_\beta$  is an even function and that for  $\beta \leq \beta_c$ ,  $\sup_x c''_\beta(x) = c''_\beta(0)$ . Therefore, by the mean value theorem and Theorem 5.2,

$$\begin{aligned} \mathbb{E}_{\sigma, \tau}[\rho(X, Y)] &\leq 1 - \frac{[1 - (n-1)(2\beta K/n)c''_\beta(0)]}{n} + O\left(\frac{1}{n^2}\right) \\ &\leq \exp \left\{ -\frac{1 - 2\beta K c''_\beta(0)}{n} + O\left(\frac{1}{n^2}\right) \right\} \\ &= \exp \left\{ \frac{1}{n} \left( \frac{K_c^{(2)}(\beta) - K}{K_c^{(2)}(\beta)} \right) + O\left(\frac{1}{n^2}\right) \right\} \\ &< e^{-\alpha/n} \end{aligned}$$

for any  $\alpha \in \left(0, \frac{K_c^{(2)}(\beta) - K}{K_c^{(2)}(\beta)}\right)$  and  $n$  sufficiently large. Thus, for  $K < K_c^{(2)}(\beta)$ , we can apply Theorem 5.10, where the diameter of the configuration space of the BC model  $\Omega^n$  is  $n$ , to complete the proof.  $\square$

**5.5. Aggregate Path Coupling in the Discontinuous Phase Transition Region.** Here we consider the region  $\beta > \beta_c$ , where the mean-field Blume-Capel model undergoes a first-order discontinuous phase transition. In this region, the function  $c'_\beta(x)$  which determines whether the mean coupling distance contracts (Corollary 5.8) is no longer strictly concave for  $x > 0$  (Theorem 5.9). See the second graph in Figure 9 for reference. We will show that rapid mixing occurs whenever  $K < K_1(\beta)$  where  $K_1(\beta)$  is the metastable critical value defined in subsection 5.1 and depicted in Figure 7.

As shown in Section 5.3, in order to apply the standard path coupling technique of Theorem 5.10, we need the inequality (17) to hold for all values of  $S_n(\sigma)$  and thus  $\sup_{[-1,1]} c''_\beta(x) < \frac{1}{2\beta K}$ . However since  $\sup_{[-1,1]} c''_\beta(x) > \frac{1}{2\beta K_1(\beta)}$ , the condition  $K < K_1(\beta)$  is not sufficient for the contraction of the mean coupling distance between *all pairs* of neighboring states which is required to prove rapid mixing using the standard path coupling technique stated in Theorem 5.10.

In order to prove rapid mixing in the region where  $\beta > \beta_c$  and  $K < K_1(\beta)$ , we take advantage of the result in Theorem 5.4 which states the weak convergence of the magnetization  $S_n/n$  to a point-mass at the origin. Thus, in the coupling of the dynamics, the magnetization of the process that starts at equilibrium will stay mainly near the origin. As a result, for two starting configurations  $\sigma$  and  $\tau$ , one of which has near-zero magnetization ( $S_n(\sigma)/n \approx 0$ ), the mean coupling distance of a coupling starting in these configurations will be the aggregate of the mean coupling distances between neighboring states along a minimal path connecting the two configurations. Although not all pairs of neighbors in the path will contract, we show that in the *aggregate*, contraction between the two configurations still holds.

In the next lemma we prove contraction of the mean coupling distance in the aggregate and then the rapid mixing result for the mean-field Blume-Capel model is proved in the theorem following the lemma by applying the new aggregate path coupling method.

**Lemma 5.12.** *Let  $(X, Y)$  be a coupling of one step of the Glauber dynamics of the BC model that begin in configurations  $\sigma$  and  $\tau$ , not necessarily neighbors with respect to the path metric  $\rho$  defined in (32). Suppose  $\beta > \beta_c$  and  $K < K_1(\beta)$ . Then for any  $\alpha \in \left(0, \frac{K_1(\beta) - K}{K_1(\beta)}\right)$  there exists an  $\varepsilon > 0$  such that, asymptotically as  $n \rightarrow \infty$ ,*

$$(19) \quad \mathbb{E}_{\sigma, \tau}[\rho(X, Y)] \leq e^{-\alpha/n} \rho(\sigma, \tau)$$

whenever  $|S_n(\sigma)| < \varepsilon n$ .

*Proof.* Observe that for  $\beta > \beta_c$  and  $K < K_1(\beta)$ ,

$$|c'_\beta(x)| \leq \frac{|x|}{2\beta K_1(\beta)} \quad \text{for all } x$$

We will show that for a given  $\alpha' \in \left(\frac{1}{2\beta K_1(\beta)}, \frac{1-\alpha}{2\beta K}\right)$ , there exists  $\varepsilon > 0$  such that

$$(20) \quad c'_\beta(x) - c'_\beta(x_0) \leq \alpha'(x - x_0) \quad \text{whenever } |x_0| < \varepsilon$$

as  $c'_\beta(x)$  is a continuously differentiable increasing odd function and  $c'_\beta(0) = 0$ .

In order to show (20), observe that  $c''_\beta(0) = \frac{1}{2\beta K_e^{(2)}(\beta)} < \frac{1}{2\beta K_1(\beta)}$ , and since  $c''_\beta$  is continuous, there exists a  $\delta > 0$  such that

$$c''_\beta(x) < \alpha' \quad \text{whenever } |x| < \delta$$

The mean value theorem implies that

$$c'_\beta(x) - c'_\beta(x_0) < \alpha'(x - x_0) \quad \text{for all } x_0, x \in (-\delta, \delta)$$

Now, let  $\varepsilon = \frac{\alpha' - 1/(2\beta K_1(\beta))}{\alpha' + 1/(2\beta K_1(\beta))} \delta < \delta$ . Then for any  $|x_0| < \varepsilon$  and  $|x| \geq \delta$ ,

$$|c'_\beta(x) - c'_\beta(x_0)| \leq \frac{|x| + |x_0|}{2\beta K_1(\beta)} \leq \frac{(1 + \varepsilon/\delta)|x|}{2\beta K_1(\beta)} = \frac{|x - x_0|}{2\beta K_1(\beta)} \cdot \frac{1 + \varepsilon/\delta}{|1 - x_0/x|} \leq \frac{|x - x_0|}{2\beta K_1(\beta)} \cdot \frac{1 + \varepsilon/\delta}{1 - \varepsilon/\delta} = \alpha'|x - x_0|$$

Without loss of generality suppose that  $S_n(\sigma) < S_n(\tau)$ . Let  $(\sigma = x_0, x_1, \dots, x_r = \tau)$  be a path connecting  $\sigma$  to  $\tau$  and monotone increasing in  $\rho$  such that  $(x_{i-1}, x_i)$  are neighboring configurations. Here  $r = \rho(\sigma, \tau)$ . Then by Corollary 5.8 of Lemma 5.7 and (20), we have for  $|S_n(\sigma)| < \varepsilon n$  and asymptotically as  $n \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{E}_{\sigma, \tau}[\rho(X, Y)] &\leq \sum_{i=1}^r \mathbb{E}_{x_{i-1}, x_i}[\rho(X_{i-1}, X_i)] \\ &= \frac{(n-1)}{n} \rho(\sigma, \tau) + \frac{(n-1)}{n} \left[ c'_\beta \left( \frac{2\beta K}{n} S_n(\tau) \right) - c'_\beta \left( \frac{2\beta K}{n} S_n(\sigma) \right) \right] + \rho(\sigma, \tau) \cdot O\left(\frac{1}{n^2}\right) \\ &\leq \frac{(n-1)}{n} \rho(\sigma, \tau) + \frac{(n-1)}{n} (S_n(\tau) - S_n(\sigma)) \frac{2\beta K \alpha'}{n} + \rho(\sigma, \tau) \cdot O\left(\frac{1}{n^2}\right) \\ &\leq \rho(\sigma, \tau) \left[ 1 - \left( \frac{1 - 2\beta K \alpha'}{n} \right) + O\left(\frac{1}{n^2}\right) \right] \\ &\leq e^{-\alpha/n} \rho(\sigma, \tau) \end{aligned}$$

This completes the proof.  $\square$

**Theorem 5.13.** *Let  $t_{\text{mix}}(\varepsilon)$  be the mixing time for the Glauber dynamics of the mean-field Blume-Capel model on  $n$  vertices and  $K_1(\beta)$  be the metastable critical point. Then, for  $\beta > \beta_c$  and  $K < K_1(\beta)$ ,*

$$t_{\text{mix}}(\varepsilon) \leq \frac{n}{\alpha} (\log n + \log(2/\varepsilon))$$

for any  $\alpha \in \left(0, \frac{K_1(\beta) - K}{K_1(\beta)}\right)$  and  $n$  sufficiently large.

*Proof.* Let  $(X_t, Y_t)$  be a coupling of the Glauber dynamics of the BC model such that  $Y_0 \stackrel{\text{dist}}{=} P_{n,\beta,K}$ , the stationary distribution. For a given  $\alpha \in \left(0, \frac{K_1(\beta)-K}{K_1(\beta)}\right)$ , let  $\varepsilon$  be as in Lemma 5.12. For sufficiently large  $n$ ,

$$\begin{aligned}
\|P^t(X_0, \cdot) - P_{n,\beta,K}\|_{TV} &\leq P\{X_t \neq Y_t\} \\
&= P\{\rho(X_t, Y_t) \geq 1\} \\
&\leq \mathbb{E}[\rho(X_t, Y_t)] \\
&= \mathbb{E}[\mathbb{E}[\rho(X_t, Y_t) \mid X_{t-1}, Y_{t-1}]] \\
&\leq \mathbb{E}[\mathbb{E}[\rho(X_t, Y_t) \mid X_{t-1}, Y_{t-1}] \mid |S_n(Y_{t-1})| < \varepsilon n] \cdot P\{|S_n(Y_{t-1})| < \varepsilon n\} \\
&\quad + nP\{|S_n(Y_{t-1})| \geq \varepsilon n\}
\end{aligned}$$

By iterating (19), it follows that

$$\begin{aligned}
\|P^t(X_0, \cdot) - P_{n,\beta,K}\|_{TV} &\leq e^{-\alpha/n} \mathbb{E}[\rho(X_{t-1}, Y_{t-1}) \mid |S_n(Y_{t-1})| < \varepsilon n] \cdot P\{|S_n(Y_{t-1})| < \varepsilon n\} \\
&\quad + nP\{|S_n(Y_{t-1})| \geq \varepsilon n\} \\
&\leq e^{-\alpha/n} \mathbb{E}[\rho(X_{t-1}, Y_{t-1})] + nP\{|S_n(Y_{t-1})| \geq \varepsilon n\} \\
&\quad \vdots \\
&\leq e^{-\alpha t/n} \mathbb{E}[\rho(X_0, Y_0)] + n \sum_{s=0}^{t-1} P\{|S_n(Y_s)| \geq \varepsilon n\} \\
&= e^{-\alpha t/n} \mathbb{E}[\rho(X_0, Y_0)] + ntP_{n,\beta,K}\{|S_n/n| \geq \varepsilon\} \\
&\leq ne^{-\alpha t/n} + ntP_{n,\beta,K}\{|S_n/n| \geq \varepsilon\}
\end{aligned}$$

We recall the result in Theorem 5.4 that for  $\beta > \beta_c$  and  $K < K_1(\beta)$

$$P_{n,\beta,K}\{S_n/n \in dx\} \implies \delta_0 \quad \text{as } n \rightarrow \infty.$$

Moreover, for any  $\gamma > 1$  and  $n$  sufficiently large, the LDP stated in Theorem 5.1 implies that

$$\begin{aligned}
\|P^t(X_0, \cdot) - P_{n,\beta,K}\|_{TV} &\leq ne^{-\alpha t/n} + ntP_{n,\beta,K}\{|S_n/n| \geq \varepsilon\} \\
&< ne^{-\alpha t/n} + tne^{-\frac{n}{\gamma}I_{\beta,K}(\varepsilon)}
\end{aligned}$$

For  $t = \frac{n}{\alpha}(\log n + \log(2/\varepsilon))$ , the above right-hand side converges to  $\varepsilon/2$  as  $n \rightarrow \infty$ .  $\square$

**5.6. Slow Mixing.** In [26], the slow mixing region of the parameter space was determined for the mean-field Blume-Capel model. Since the method used to prove the slow mixing, called the bottleneck ratio or Cheeger constant method, is not a coupling method, we simply state the result for completeness.

**Theorem 5.14.** *Let  $t_{\text{mix}} = t_{\text{mix}}(1/4)$  be the mixing time for the Glauber dynamics of the mean-field Blume-Capel model on  $n$  vertices. For (a)  $\beta \leq \beta_c$  and  $K > K_c^{(2)}(\beta)$ , and (b)  $\beta > \beta_c$  and  $K > K_1(\beta)$ , there exists a positive constant  $b$  and a strictly positive function  $r(\beta, K)$  such that*

$$t_{\text{mix}} \geq be^{r(\beta,K)n}$$



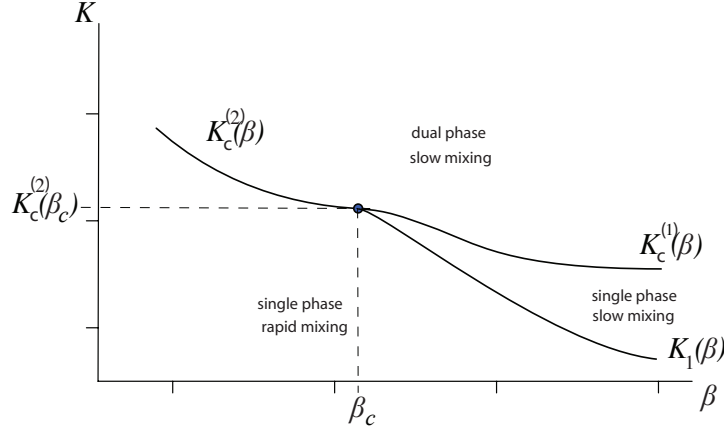


FIGURE 10. Mixing times and equilibrium phase transition structure of the mean-field Blume-Capel model

We summarize the mixing time results for the mean-field Blume-Capel model and its relationship to the model's thermodynamic phase transition structure in Figure 10. As shown in the figure, in the second-order, continuous phase transition region ( $\beta \leq \beta_c$ ) for the BC model, the mixing time transition coincides with the equilibrium phase transition. This is consistent with other models that exhibit this type of phase transition. However, in the first-order, discontinuous phase transition region ( $\beta > \beta_c$ ) the mixing time transition occurs below the equilibrium phase transition at the metastable critical value.

## 6. AGGREGATE PATH COUPLING FOR GENERAL CLASS OF GIBBS ENSEMBLES

In this section, we extend the aggregate path coupling technique derived in the previous section for the Blume-Capel model to a large class of statistical mechanical models that is disjoint from the mean-field Blume-Capel model. The aggregate path coupling method presented here extends the classical path coupling method for Gibbs ensembles in two directions. First, we consider macroscopic quantities in higher dimensions and find a monotone contraction path by considering a related variational problem in the continuous space. We also do not require the monotone path to be a nearest-neighbor path. In fact, in most situations we consider, a nearest-neighbor path will not work for proving contraction. Second, the aggregation of the mean path distance along a monotone path is shown to contract for some but not all pairs of configurations. Yet, we use measure concentration and large deviation principle to show that showing contraction for pairs of configurations, where at least one of them is close enough to the equilibrium, is sufficient for establishing rapid mixing.

Our main result is general enough to be applied to statistical mechanical models that undergo both types of phase transitions and to models whose macroscopic quantity are in higher dimensions. Moreover, despite the generality, the application of our results requires straightforward conditions that we illustrate in Section 7. This is a significant simplification for proving rapid mixing for statistical mechanical models, especially those that undergo first-order, discontinuous phase transitions. Lastly, our results also provide a link between measure concentration of the stationary distribution and rapid mixing of the corresponding dynamics for this class of statistical mechanical models. This idea has been previously studied in [31] where the main result showed that rapid mixing implied measure concentration defined in terms of Lipschitz functions. In our work, we prove a type of converse where measure concentration, in terms of a large deviation principle, implies rapid mixing.

**6.1. Class of Gibbs Ensembles.** We begin by defining the general class of statistical mechanical spin models for which our results can be applied. In Section 7, we illustrate the application of our main result for the generalized Curie-Weiss-Potts model for which the mixing times has not been previously obtained.

Let  $q$  be a fixed integer and define  $\Lambda = \{e^1, e^2, \dots, e^q\}$ , where  $e^k$  are the  $q$  standard basis vectors of  $\mathbb{R}^q$ . A *configuration* of the model has the form  $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Lambda^n$ . We will consider a configuration on a graph with  $n$  vertices and let  $X_i(\omega) = \omega_i$  be the *spin* at vertex  $i$ . The random variables  $X_i$ 's for  $i = 1, 2, \dots, n$  are independent and identically distributed with common distribution  $\rho$ .

In terms of the microscopic quantities, the spins at each vertex, the relevant macroscopic quantity is the *magnetization vector* (a.k.a empirical measure or proportion vector)

$$(21) \quad L_n(\omega) = (L_{n,1}(\omega), L_{n,2}(\omega), \dots, L_{n,q}(\omega))$$

where the  $k$ th component is defined by

$$L_{n,k}(\omega) = \frac{1}{n} \sum_{i=1}^n \delta(\omega_i, e^k)$$

which yields the proportion of spins in configuration  $\omega$  that take on the value  $e^k$ . The magnetization vector  $L_n$  takes values in the set of probability vectors

$$(22) \quad \mathcal{P}_n = \left\{ \frac{n_k}{n} : \text{each } n_k \in \{0, 1, \dots, n\} \text{ and } \sum_{k=1}^q n_k = n \right\}$$

inside the continuous simplex

$$\mathcal{P} = \left\{ \nu \in \mathbb{R}^q : \nu = (\nu_1, \nu_2, \dots, \nu_q), \text{ each } \nu_k \geq 0, \sum_{k=1}^q \nu_k = 1 \right\}.$$

**Remark 6.1.** For  $q = 2$ , the empirical measure  $L_n$  yields the empirical mean  $S_n(\omega)/n$  where  $S_n(\omega) = \sum_{i=1}^n \omega_i$ . Therefore, the class of models considered in this paper includes those where the relevant macroscopic quantity is the empirical mean, like the Curie-Weiss (mean-field Ising) model.

As discussed in section 3, statistical mechanical models are defined in terms of the Hamiltonian function, denoted by  $H_n(\omega)$ , which encodes the interactions of the individual spins and the total energy of a configuration. The link between the microscopic interactions to the macroscopic quantity, in this case  $L_n(\omega)$ , is the interaction representation function, which we define again for convenience below.

**Definition 6.2.** For  $z \in \mathbb{R}^q$ , we define the **interaction representation function**, denoted by  $H(z)$ , to be a differentiable function satisfying

$$H_n(\omega) = nH(L_n(\omega))$$

Throughout the paper we suppose the interaction representation function  $H(z)$  is a finite concave  $\mathcal{C}^3(\mathbb{R}^q)$  function that has the form

$$H(z) = H_1(z_1) + H_2(z_2) + \dots + H_q(z_q)$$

For example, for the Curie-Weiss-Potts (CWP) model [11],

$$H(z) = -\frac{1}{2}\langle z, z \rangle = -\frac{1}{2}z_1^2 - \frac{1}{2}z_2^2 - \dots - \frac{1}{2}z_q^2.$$

The class Gibbs measures or Gibbs ensemble considered in this section is defined by

$$(23) \quad P_{n,\beta}(B) = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta H_n(\omega)\} dP_n = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta n H(L_n(\omega))\} dP_n$$

where  $P_n$  is the product measure with identical marginals  $\rho$  and  $Z_n(\beta) = \int_{\Lambda^n} \exp\{-\beta H_n(\omega)\} dP_n$  is the **partition function**. The positive parameter  $\beta$  represents the inverse temperature of the external heat bath.

**Remark 6.3.** To simplify the presentation, we take  $\Lambda = \{e^1, e^2, \dots, e^q\}$ , where  $e^k$  are the  $q$  standard basis vectors of  $\mathbb{R}^q$ . But our analysis has a straight-forward generalization to the case where  $\Lambda = \{\theta^1, \theta^2, \dots, \theta^q\}$ , where  $\theta^k$  is any basis of  $\mathbb{R}^q$ . In this case, the product measure  $P_n$  would have identical one-dimensional marginals equal to

$$\bar{\rho} = \frac{1}{q} \sum_{i=1}^q \delta_{\theta^i}$$

An important tool we use to prove rapid mixing of the Glauber dynamics that converge to the Gibbs ensemble above is the large deviation principle of the empirical measure with respect to the Gibbs ensemble. This measure concentration is precisely what drives the rapid mixing. The large deviation principle for our class of Gibbs ensembles  $P_{n,\beta}$  is presented next.

**6.2. Large Deviations.** By Sanov's Theorem, the empirical measure  $L_n$  satisfies the large deviation principle (LDP) with respect to the product measure  $P_n$  with identical marginals  $\rho$  and the rate function is given by the **relative entropy**

$$R(\nu|\rho) = \sum_{k=1}^q \nu_k \log \left( \frac{\nu_k}{\rho_k} \right)$$

for  $\nu \in \mathcal{P}$ . Theorem 2.4 of [18] yields the following result for the Gibbs measures (23).

**Theorem 6.4.** *The empirical measure  $L_n$  satisfies the LDP with respect to the Gibbs measure  $P_{n,\beta}$  with rate function*

$$I_\beta(z) = R(z|\rho) + \beta H(z) - \inf_t \{R(t|\rho) + \beta H(t)\}.$$

As discussed in section 4, the LDP upper bound stated in the previous theorem yields the following natural definition of **equilibrium macrostates** of the model.

$$(24) \quad \mathcal{E}_\beta := \{\nu \in \mathcal{P} : \nu \text{ minimizes } R(\nu|\rho) + \beta H(\nu)\}$$

For our main result, we assume that there exists a positive interval  $B$  such that for all  $\beta \in B$ ,  $\mathcal{E}_\beta$  consists of a single state  $z_\beta$ . We refer to this interval  $B$  as the single phase region.

Again from the LDP upper bound, when  $\beta$  lies in the single phase region, we get

$$(25) \quad P_{n,\beta}(L_n \in dx) \implies \delta_{z_\beta} \quad \text{as} \quad n \rightarrow \infty.$$

The above asymptotic behavior will play a key role in obtaining a rapid mixing time rate for the Glauber dynamics corresponding to the Gibbs measures (23).

An important function in our work is the free energy functional defined below. It is defined in terms of the interaction representation function  $H$  and the logarithmic moment generating function of a single spin; specifically, for  $z \in \mathbb{R}^q$  and  $\rho$  equal to the uniform distribution, the **logarithmic moment generating function** of  $X_1$ , the spin at vertex 1, is defined by

$$(26) \quad \Gamma(z) = \log \left( \frac{1}{q} \sum_{k=1}^q \exp\{z_k\} \right).$$

**Definition 6.5.** *The free energy functional for the Gibbs ensemble  $P_{n,\beta}$  is defined as*

$$(27) \quad G_\beta(z) = \beta(-H)^*(-\nabla H(z)) - \Gamma(-\beta \nabla H(z))$$

where for a finite, differentiable, convex function  $F$  on  $\mathbb{R}^q$ ,  $F^*$  denotes its Legendre-Fenchel transform defined by

$$F^*(z) = \sup_{x \in \mathbb{R}^q} \{\langle x, z \rangle - F(x)\}$$

The following lemma yields an alternative formulation of the set of equilibrium macrostates of the Gibbs ensemble in terms of the free energy functional. The proof is a straightforward generalization of Theorem A.1 in [10].

**Lemma 6.6.** *Suppose  $H$  is finite, differentiable, and concave. Then*

$$\inf_{z \in \mathcal{P}} \{R(z|\rho) + \beta H(z)\} = \inf_{z \in \mathbb{R}^q} \{G_\beta(z)\}$$

Moreover,  $z_0 \in \mathcal{P}$  is a minimizer of  $R(z|\rho) + \beta H(z)$  if and only if  $z_0$  is a minimizer of  $G_\beta(z)$ .

Therefore, the set of equilibrium macrostates can be expressed in terms of the free energy functional as

$$(28) \quad \mathcal{E}_\beta = \{z \in \mathcal{P} : z \text{ minimizes } G_\beta(z)\}$$

As mentioned above, we consider only the single phase region of the Gibbs ensemble; i.e. values of  $\beta$  where  $G_\beta(z)$  has a unique global minimum. For example, for the Curie-Weiss-Potts model [10], the single phase region are values of  $\beta$  such that  $0 < \beta < \beta_c := (2(q-1)/(q-2)) \log(q-1)$ .

At this critical value  $\beta_c$ , the model undergoes a first-order, discontinuous phase transition in which the single phase changes to a multiple phase discontinuously.

As we will show, the geometry of the free energy functional  $G_\beta$  not only determines the equilibrium behavior of the Gibbs ensembles but it also yields the condition for rapid mixing of the corresponding Glauber dynamics.

**6.3. Glauber Dynamics.** On the configuration space  $\Lambda^n$ , we define the Glauber dynamics, defined in general in subsection 5.2, for the class of Gibbs ensembles  $P_{n,\beta}$  defined in (23). These dynamics yield a reversible Markov chain  $X^t$  with stationary distribution being the Gibbs ensemble  $P_{n,\beta}$ .

- (i) Select a vertex  $i$  uniformly,
- (ii) Update the spin at vertex  $i$  according to the distribution  $P_{n,\beta}$ , conditioned on the event that the spins at all vertices not equal to  $i$  remain unchanged.

For a given configuration  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ , denote by  $\sigma_{i,e^k}$  the configuration that agrees with  $\sigma$  at all vertices  $j \neq i$  and the spin at the vertex  $i$  is  $e^k$ ; i.e.

$$\sigma_{i,e^k} = (\sigma_1, \sigma_2, \dots, \sigma_{i-1}, e^k, \sigma_{i+1}, \dots, \sigma_n)$$

Then if the current configuration is  $\sigma$  and vertex  $i$  is selected, the probability the spin at  $i$  is updated to  $e^k$ , denoted by  $P(\sigma \rightarrow \sigma_{i,e^k})$ , is equal to

$$(29) \quad P(\sigma \rightarrow \sigma_{i,e^k}) = \frac{\exp\{-\beta n H(L_n(\sigma_{i,e^k}))\}}{\sum_{\ell=1}^q \exp\{-\beta n H(L_n(\sigma_{i,e^\ell}))\}}.$$

Next, we show that the update probabilities of the Glauber dynamics above can be expressed in terms of the derivative of the logarithmic moment generating function of the individual spins  $\Gamma$  defined in (26). The partial derivative of  $\Gamma$  in the direction of  $e^\ell$  has the form

$$[\partial_\ell \Gamma](z) = \frac{\exp\{z_\ell\}}{\sum_{k=1}^q \exp\{z_k\}}$$

We introduce the following function that plays the key role in our analysis.

$$(30) \quad g_\ell^{H,\beta}(z) = [\partial_\ell \Gamma](-\beta \nabla H(z)) = \frac{\exp(-\beta [\partial_\ell H](z))}{\sum_{k=1}^q \exp(-\beta [\partial_k H](z))}.$$

Denote

$$(31) \quad g^{H,\beta}(z) := (g_1^{H,\beta}(z), \dots, g_q^{H,\beta}(z)).$$

Note that  $g^{H,\beta}(z)$  maps the simplex

$$\mathcal{P} = \left\{ \nu \in \mathbb{R}^q : \nu = (\nu_1, \nu_2, \dots, \nu_q), \text{ each } \nu_k \geq 0, \sum_{k=1}^q \nu_k = 1 \right\}$$

into itself and it can be expressed in terms of the free energy functional  $G_\beta$  defined in (27) by

$$\nabla G_\beta(z) = \beta[\nabla(-H)^*(-\nabla H(z)) - g^{H,\beta}(z)]$$

**Lemma 6.7.** *Let  $P(\sigma \rightarrow \sigma_{i,e^k})$  be the Glauber dynamics update probabilities given in (29). Then, for any  $k \in \{1, 2, \dots, q\}$ ,*

$$P(\sigma \rightarrow \sigma_{i,e^k}) = [\partial_k \Gamma] \left( -\beta \nabla H(L_n(\sigma)) - \frac{\beta}{2n} \mathcal{Q}H(L_n(\sigma)) + \frac{\beta}{n} \left\langle \sigma_i, \mathcal{Q}H(L_n(\sigma)) \right\rangle \sigma_i \right) + O\left(\frac{1}{n^2}\right),$$

where  $\mathcal{Q}$  is the following linear operator:

$$\mathcal{Q}F(z) := (\partial_1^2 F(z), \partial_2^2 F(z), \dots, \partial_q^2 F(z)),$$

for any  $F : \mathbb{R}^q \rightarrow \mathbb{R}$  in  $\mathcal{C}^2$ .

*Proof.* Suppose  $\sigma_i = e^m$ . By Taylor's theorem, for any  $k \neq m$ , we have

$$\begin{aligned} H(L_n(\sigma_{i,e^k})) &= H(L_n(\sigma)) + H_m(L_{n,m}(\sigma) - 1/n) - H_m(L_{n,m}(\sigma)) \\ &\quad + H_k(L_{n,k}(\sigma) + 1/n) - H_k(L_{n,k}(\sigma)) \\ &= H(L_n(\sigma)) + \frac{1}{n} [\partial_k H(L_n(\sigma)) - \partial_m H(L_n(\sigma))] \\ &\quad + \frac{1}{2n^2} [\partial_k^2 H(L_n(\sigma)) + \partial_m^2 H(L_n(\sigma))] + O\left(\frac{1}{n^3}\right). \end{aligned}$$

Now, if  $k = m$ ,

$$\begin{aligned} H(L_n(\sigma_{i,e^k})) &= H(L_n(\sigma)) \\ &= H(L_n(\sigma)) + \frac{1}{n} [\partial_k H(L_n(\sigma)) - \partial_m H(L_n(\sigma))] \\ &\quad + \frac{1}{2n^2} [-\partial_k^2 H(L_n(\sigma)) + \partial_m^2 H(L_n(\sigma))]. \end{aligned}$$

This implies that the transition probability (29) has the form

$$P(\sigma \rightarrow \sigma_{i,e^k}) = [\partial_k \Gamma] \left( -\beta \nabla H(L_n(\sigma)) - \frac{\beta}{2n} \mathcal{Q}H(L_n(\sigma)) + \frac{\beta}{n} \partial_m^2 H(L_n(\sigma)) e^m \right) + O\left(\frac{1}{n^2}\right)$$

as  $\exp\left\{O\left(\frac{1}{n^2}\right)\right\} = 1 + O\left(\frac{1}{n^2}\right)$ .  $\square$

The above Lemma 6.7 can be restated as follows using Taylor expansion.

**Corollary 6.8.** *Let  $P(\sigma \rightarrow \sigma_{i,e^k})$  be the Glauber dynamics update probabilities given in (29). Then, for any  $k \in \{1, 2, \dots, q\}$ ,*

$$P(\sigma \rightarrow \sigma_{i,e^k}) = g_k^{H,\beta}(L_n(\sigma)) + \frac{\beta}{n} \varphi_{k,\sigma_i}^{H,\beta}(L_n(\sigma)) + O\left(\frac{1}{n^2}\right),$$

where

$$\varphi_{k,e^r}^{H,\beta}(z) := -\frac{1}{2} \left\langle \mathcal{Q}H(z), [\nabla \partial_\ell \Gamma](-\beta \nabla H(z)) \right\rangle + \left\langle e^r, \mathcal{Q}H(z) \right\rangle \left\langle e^r, [\nabla \partial_\ell \Gamma](-\beta \nabla H(z)) \right\rangle.$$

In the next section, we define the specific coupling used to bound the mixing time of the Glauber dynamics for the class of Gibbs ensembles defined in subsection 6.1.

**6.4. Coupling of Glauber Dynamics.** We begin by defining a metric on the configuration space  $\Lambda^n$ . For two configurations  $\sigma$  and  $\tau$  in  $\Lambda^n$ , define

$$(32) \quad d(\sigma, \tau) = \sum_{j=1}^n 1\{\sigma_j \neq \tau_j\}$$

which yields the number of vertices at which the two configurations differ.

Let  $X^t$  and  $Y^t$  be two copies of the Glauber dynamics. Here, we use the standard greedy coupling of  $X^t$  and  $Y^t$ . At each time step a vertex is selected at random, uniformly from the  $n$  vertices. Suppose  $X^t = \sigma$ ,  $Y^t = \tau$ , and the vertex selected is denoted by  $j$ . Next, we erase the spin at location  $j$  in both processes, and replace it with a new one according to the following update probabilities. For all  $\ell = 1, 2, \dots, q$ , define

$$p_\ell = P(\sigma \rightarrow \sigma_{j,e^\ell}) \quad \text{and} \quad q_\ell = P(\tau \rightarrow \tau_{j,e^\ell})$$

and let

$$P_\ell = \min\{p_\ell, q_\ell\} \quad \text{and} \quad P = \sum_{\ell=1}^q P_\ell.$$

Now, let  $B$  be a Bernoulli random variable with probability of success  $P$ . If  $B = 1$ , we update the two chains equally with the following probabilities

$$P(X_j^{t+1} = e^\ell, Y_j^{t+1} = e^\ell \mid B = 1) = \frac{P_\ell}{P}$$

for  $\ell = 1, 2, \dots, q$ . On the other hand, if  $B = 0$ , we update the chains differently according to the following probabilities

$$P(X_j^{t+1} = e^\ell, Y_j^{t+1} = e^m \mid B = 0) = \frac{p_\ell - P_\ell}{1 - P} \cdot \frac{q_m - P_m}{1 - P}$$

for all pairs  $\ell \neq m$ . Then the total probability that the two chains update the same is equal to  $P$  and the total probability that the chains update differently is equal to  $1 - P$ .

Observe that once  $X^t = Y^t$ , the processes remain matched (coupled) for the rest of the time. In the coupling literature, the time

$$\min\{t \geq 0 : X^t = Y^t\}$$

is referred to as the *coupling time*.

As discussed in section 2, the mean coupling distance  $\mathbb{E}[d(X^t, Y^t)]$  is tied to the total variation distance via the following inequality:

$$(33) \quad \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \leq P(X^t \neq Y^t) \leq \mathbb{E}[d(X^t, Y^t)]$$

The above inequality implies that the order of the mean coupling time is an upper bound on the order of the mixing time. See [30] and [29] for details on coupling and coupling inequalities.

**6.5. Mean Coupling Distance.** Fix  $\varepsilon > 0$ . Consider two configurations  $\sigma$  and  $\tau$  such that

$$d(\sigma, \tau) = d,$$

where  $d(\sigma, \tau) \in \mathbb{N}$  is the metric defined in (32) and  $\varepsilon \leq \|L_n(\sigma) - L_n(\tau)\|_1 < 2\varepsilon$ .

Let  $\mathcal{I} = \{i_1, \dots, i_d\}$  be the set of vertices at which the spin values of the two configurations  $\sigma$  and  $\tau$  disagree. Define  $\kappa(e^\ell)$  to be the probability that the coupled processes update differently when the chosen vertex  $j \notin \mathcal{I}$  has spin  $e^\ell$ . If the chosen vertex  $j$  is such that  $\sigma_j = \tau_j = e^\ell$ , then by Corollary 6.8 of Lemma 6.7,

$$\begin{aligned} \kappa(e^\ell) &:= \frac{1}{2} \sum_{k=1}^q \left| P(\sigma \rightarrow \sigma_{j,e^k}) - P(\tau \rightarrow \tau_{j,e^k}) \right| \\ &= \frac{1}{2} \sum_{k=1}^q \left| \left( g_k^{H,\beta}(L_n(\sigma)) + \frac{\beta}{n} \varphi_{k,e^\ell}^{H,\beta}(L_n(\sigma)) \right) - \left( g_k^{H,\beta}(L_n(\tau)) + \frac{\beta}{n} \varphi_{k,e^\ell}^{H,\beta}(L_n(\tau)) \right) \right| + O\left(\frac{1}{n^2}\right) \\ (34) \quad &= \frac{1}{2} \sum_{k=1}^q \left| g_k^{H,\beta}(L_n(\sigma)) - g_k^{H,\beta}(L_n(\tau)) \right| + O\left(\frac{\varepsilon}{n} + \frac{1}{n^2}\right). \end{aligned}$$

Next, we observe that for any  $\mathcal{C}^2$  function  $f : \mathcal{P} \rightarrow \mathbb{R}$ , there exists  $C > 0$  such that

$$(35) \quad \left| f(z') - f(z) - \langle z' - z, \nabla f(z) \rangle \right| < C\varepsilon^2$$

for all  $z, z' \in \mathcal{P}$  satisfying  $\varepsilon \leq \|z' - z\|_1 < 2\varepsilon$ .

Therefore for  $n$  large enough, there exists  $C' > 0$  such that

$$(36) \quad \left| \kappa(e^\ell) - \frac{1}{2} \sum_{k=1}^q \left| \langle L_n(\tau) - L_n(\sigma), \nabla g_k^{H,\beta}(L_n(\sigma)) \rangle \right| \right| < C'\varepsilon^2.$$

The above result holds regardless of the value of  $\ell \in \{1, 2, \dots, q\}$ .

Similarly, when the chosen vertex  $j \in \mathcal{I}$ , the probability of not coupling at  $j$  satisfies (36).

We conclude that in terms of  $\kappa_{\sigma,\tau} := \frac{1}{2} \sum_{k=1}^q \left| \langle L_n(\tau) - L_n(\sigma), \nabla g_k^{H,\beta}(L_n(\sigma)) \rangle \right|$ , the mean distance between a coupling of the Glauber dynamics starting in  $\sigma$  and  $\tau$  with  $d(\sigma, \tau) = d$  after one step has the form

$$\begin{aligned} \mathbb{E}_{\sigma,\tau}[d(X, Y)] &\leq d - \frac{d}{n}(1 - \kappa_{\sigma,\tau}) + \frac{n-d}{n}\kappa_{\sigma,\tau} + c\varepsilon^2 \\ (37) \quad &= d \cdot \left[ 1 - \frac{1}{n} \left( 1 - \frac{\kappa_{\sigma,\tau} + c\varepsilon^2}{d/n} \right) \right] \end{aligned}$$

for a fixed  $c > 0$  and all  $\varepsilon$  small enough.



**6.6. Aggregate Path Coupling.** In the previous section, we derived the form of the mean distance between a coupling of the Glauber dynamics starting in two configurations whose distance is bounded. We next derive the form of the mean coupling distance of a coupling starting in two configurations that are connected by a path of configurations where the distance between successive configurations are bounded.

**Definition 6.9.** Let  $\sigma$  and  $\tau$  be configurations in  $\Lambda^n$ . We say that a path  $\pi$  connecting configurations  $\sigma$  and  $\tau$  denoted by

$$\pi : \sigma = x_0, x_1, \dots, x_r = \tau,$$

is a **monotone path** if

- (i)  $\sum_{i=1}^r d(x_{i-1}, x_i) = d(\sigma, \tau)$
- (ii) for each  $k = 1, 2, \dots, q$ , the  $k$ th coordinate of  $L_n(x_i)$ ,  $L_{n,k}(x_i)$  is monotonic as  $i$  increases from 0 to  $r$ ;

Observe that here the points  $x_i$  on the path are not required to be nearest-neighbors.

A straightforward property of monotone paths is that

$$\sum_{i=1}^r \sum_{k=1}^q L_{n,k}(x_i) - L_{n,k}(x_{i-1}) = L_n(\sigma) - L_n(\tau)$$

Another straightforward observation is that for any given path

$$L_n(\sigma) = z_0, z_1, \dots, z_r = L_n(\tau)$$

in  $\mathcal{P}_n$ , monotone in each coordinate, with  $\|z_i - z_{i-1}\|_1 > 0$  for all  $i \in \{1, 2, \dots, r\}$ , there exists a monotone path

$$\pi : \sigma = x_0, x_1, \dots, x_r = \tau$$

such that  $L_n(x_i) = z_i$  for each  $i$ .

Let  $\pi : \sigma = x_0, x_1, \dots, x_r = \tau$  be a monotone path connecting configurations  $\sigma$  and  $\tau$  such that  $\varepsilon \leq \|L_n(x_i) - L_n(x_{i-1})\|_1 < 2\varepsilon$  for all  $i = 1, \dots, r$ . Equation (37) implies the following bound on the mean distance between a coupling of the Glauber dynamics starting in configurations  $\sigma$  and  $\tau$ :

$$\begin{aligned}
& \mathbb{E}_{\sigma, \tau}[d(X, Y)] \\
& \leq \sum_{i=1}^r \mathbb{E}_{x_{i-1}, x_i}[d(X_{i-1}, X_i)] \\
& \leq \sum_{i=1}^r \left\{ d(x_{i-1}, x_i) \cdot \left[ 1 - \frac{1}{n} \left( 1 - \frac{\frac{1}{2} \sum_{k=1}^q \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right| + c\varepsilon^2}{d(x_{i-1}, x_i)/n} \right) \right] \right\} \\
& = d(\sigma, \tau) \left[ 1 - \frac{1}{n} \left( 1 - \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right| + c\varepsilon^2}{2d(\sigma, \tau)/n} \right) \right] \\
(38) \quad & \leq d(\sigma, \tau) \left[ 1 - \frac{1}{n} \left( 1 - \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right| + c\varepsilon^2}{\|L_n(\sigma) - L_n(\tau)\|_1} \right) \right],
\end{aligned}$$

as  $\sum_{i=1}^r d(x_{i-1}, x_i) = d(\sigma, \tau)$ .

From inequality (38), if there exists monotone paths between all pairs of configurations such that there is a uniform bound less than 1 on the ratio

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1}$$

then the mean coupling distance contracts which yields a bound on the mixing time via coupling inequality (33).

Although the Gibbs measure are distributions of the empirical measure  $L_n$  defined on the discrete space  $\mathcal{P}_n$ , proving contraction of the mean coupling distance is often facilitated by working in the continuous space, namely the simplex  $\mathcal{P}$ . We begin our discussion of aggregate path coupling by defining distances along paths in  $\mathcal{P}$ .

Recall the function  $g^{H, \beta}$  defined in (31) which is dependent on the Hamiltonian of the model through the interaction representation function  $H$  defined in Definition 6.2.

**Definition 6.10.** Define the **aggregate  $g$ -variation** between a pair of points  $x$  and  $z$  in  $\mathcal{P}$  along a continuous monotone (in each coordinate) path  $\rho$  to be

$$D_\rho^g(x, z) := \sum_{k=1}^q \int_\rho \left| \left\langle \nabla g_k^{H, \beta}(y), dy \right\rangle \right|$$

Define the corresponding **pseudo-distance** between a pair of points  $x$  and  $z$  in  $\mathcal{P}$  as

$$d_g(x, z) := \inf_\rho D_\rho^g(x, z),$$

where the infimum is taken over all continuous monotone paths in  $\mathcal{P}$  connecting  $x$  and  $z$ .

Notice if the monotonicity restriction is removed, the above infimum would satisfy the triangle inequality. We will need the following condition.

**Condition 6.11.** *Let  $z_\beta$  be the unique equilibrium macrostate. There exists  $\delta \in (0, 1)$  such that*

$$\frac{d_g(z, z_\beta)}{\|z - z_\beta\|_1} \leq 1 - \delta$$

for all  $z$  in  $\mathcal{P}$ .

Observe that if it is shown that  $d_g(z, z_\beta) < \|z - z_\beta\|_1$  for all  $z$  in  $\mathcal{P}$ , then by continuity the above condition is equivalent to

$$\limsup_{z \rightarrow z_\beta} \frac{d_g(z, z_\beta)}{\|z - z_\beta\|_1} < 1$$

Suppose Condition 6.11 is satisfied. Then let denote by  $\mathbf{NG}_\delta$  the family of smooth curves, monotone in each coordinate such that for each  $z \neq z_\beta$  in  $\mathcal{P}$ , there is exactly one curve  $\rho = \rho_z$  in the family  $\mathbf{NG}_\delta$  connecting  $z_\beta$  to  $z$ , and

$$\frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} \leq 1 - \delta/2.$$

Such family of smooth curves will be referred to as **neo-geodesic**.

**Condition 6.12.** *For  $\varepsilon > 0$  small enough, there exists a neo-geodesic family  $\mathbf{NG}_\delta$  such that for each  $z$  in  $\mathcal{P}$  satisfying  $\|z - z_\beta\|_1 \geq \varepsilon$ , the curve  $\rho = \rho_z$  in the family  $\mathbf{NG}_\delta$  that connects  $z_\beta$  to  $z$  satisfies*

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle z_i - z_{i-1}, \nabla g_k^{H,\beta}(z_{i-1}) \right\rangle \right|}{\|z - z_\beta\|_1} \leq 1 - \delta/3$$

for a sequence of points  $z_0 = z_\beta, z_1, \dots, z_r = z$  interpolating  $\rho$  such that

$$\varepsilon \leq \|z_i - z_{i-1}\|_1 < 2\varepsilon \quad \text{for } i = 1, 2, \dots, r.$$

It is important to observe that Condition 6.11 is often simpler to verify than Condition 6.12. Moreover, under certain simple additional prerequisites, Condition 6.11 implies Condition 6.12. For example, this is achieved if there is a uniform bound on the Cauchy curvature at every point of every curve in  $\mathbf{NG}_\delta$ . So it will be demonstrated on the example of the generalized Curie-Weiss-Potts model in section 7 that the natural way for establishing Condition 6.12 for the model is via first establishing Condition 6.11.

In addition to Condition 6.12 that will be shown to imply contraction when one of the two configurations in the coupled processes is at the equilibrium, i.e.  $L_n(\sigma) = z_\beta$ , we need a condition that will imply contraction between two configurations within a neighborhood of the equilibrium configuration. We state this assumption next.

**Condition 6.13.** *Let  $z_\beta$  be the unique equilibrium macrostate. Then,*

$$\limsup_{z \rightarrow z_\beta} \frac{\|g^{H,\beta}(z) - g^{H,\beta}(z_\beta)\|_1}{\|z - z_\beta\|_1} < 1.$$

Since  $H(z) \in \mathcal{C}^3$ , the above Condition 6.13 implies that for any  $\varepsilon > 0$  sufficiently small, there exists  $\gamma \in (0, 1)$  such that

$$\frac{\|g^{H,\beta}(z) - g^{H,\beta}(w)\|_1}{\|z - w\|_1} < 1 - \gamma$$

for all  $z$  and  $w$  in  $\mathcal{P}$  satisfying

$$\|z - z_\beta\|_1 < \varepsilon \quad \text{and} \quad \|w - z_\beta\|_1 < \varepsilon.$$

**6.7. Main Result.** As discussed in section 2, a sufficient condition for rapid mixing of the Glauber dynamics of Gibbs ensembles is contraction of the mean coupling distance  $\mathbb{E}_{\sigma,\tau}[d(X, Y)]$  between coupled processes starting in all pairs of configurations in  $\Lambda^n$ . The classical path coupling argument stated in Proposition 2.4 is a method of obtaining this contraction by only proving contraction between couplings starting in neighboring configurations. However for some classes of models (e.g. models that undergo a first-order, discontinuous phase transition) there are situations when Glauber dynamics exhibits rapid mixing, but coupled processes do not exhibit contraction between some neighboring configurations. A major strength of the **aggregate path coupling** method is that it yields a proof for rapid mixing even in those cases when contraction of the mean distance between couplings starting in all pairs of neighboring configurations does not hold.

The strategy is to take advantage of the large deviations estimates discussed in section 6.2. Recall from that section that we assume that the set of equilibrium macrostates  $\mathcal{E}_\beta$ , which can be expressed in the form given in (28), consists of a single point  $z_\beta$ . Define an *equilibrium configuration*  $\sigma_\beta$  to be a configuration such that

$$L_n(\sigma_\beta) = z_\beta = ((z_\beta)_1, (z_\beta)_2, \dots, (z_\beta)_q).$$

First we observe that in order to use the coupling inequality (33) we need to show contraction of the mean coupling distance  $\mathbb{E}_{\sigma,\tau}[d(X, Y)]$  between a Markov chain initially distributed according to the stationary probability distribution  $P_{n,\beta}$  and a Markov chain starting at any given configuration. Using large deviations we know that with high probability the former process starts near the equilibrium and stays near the equilibrium for long duration of time.

Our main result Theorem 6.15 states that once we establish contraction of the mean coupling distance between two copies of a Markov chain where one of the coupled dynamics starts near an equilibrium configuration in Lemma 6.14, then this contraction, along with the large deviations estimates of the empirical measure  $L_n$ , yields rapid mixing of the Glauber dynamics converging to the Gibbs measure.

Now, the classical path coupling relies on showing contraction along any monotone path connecting two configurations, in one time step. Here we observe that we only need to show contraction along one monotone path connecting two configurations in order to have the mean coupling distance  $\mathbb{E}_{\sigma,\tau}[d(X, Y)]$  contract in a single time step. However, finding even one monotone path with which we can show contraction in the equation (38) is not easy. The answer to

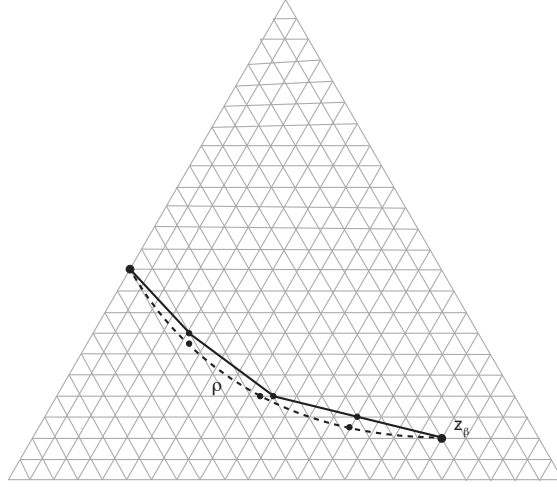


FIGURE 11. Case  $q = 3$ . Dashed curve is the continuous monotone path  $\rho$ . Solid lines represent the path  $L_n(x_0), L_n(x_1), \dots, L_n(x_r)$  in  $\mathcal{P}_n$ .

this is in finding a monotone path  $\rho$  in  $\mathcal{P}$  connecting the  $L_n$  values of the two configurations,  $\sigma$  and  $\tau$ , such that

$$\frac{\sum_{k=1}^q \int_{\rho} \left| \left\langle \nabla g_k^{H,\beta}(y), dy \right\rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1} < 1$$

Although  $\rho$  is a continuous path in continuous space  $\mathcal{P}$ , it serves as Ariadne's thread for finding a monotone path

$$\pi : \sigma = x_0, x_1, \dots, x_r = \tau$$

such that  $L_n(x_0), L_n(x_1), \dots, L_n(x_r)$  in  $\mathcal{P}_n$  are positioned along  $\rho$ , and

$$\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1})) \right\rangle \right|$$

is a Riemann sum approximating  $\sum_{k=1}^q \int_{\rho} \left| \left\langle \nabla g_k^{H,\beta}(y), dy \right\rangle \right|$ . Therefore we obtain

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1})) \right\rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1} < 1,$$

that in turn implies contraction in (38) for  $\varepsilon$  small enough and  $n$  large enough. See Figure 11.

Observe that in order for the above argument to work, we need to spread points  $L_n(x_i) \in \mathcal{P}_n$  along a continuous path  $\rho$  at intervals of fixed order  $\varepsilon$ . Thus  $\pi$  has to be **not a nearest-neighbor path** in the space of configurations, another significant deviation from the classical path coupling.

**Lemma 6.14.** *Assume Condition 6.12 and Condition 6.13. Let  $(X, Y)$  be a coupling of the Glauber dynamics as defined in Section 6.4, starting in configurations  $\sigma$  and  $\tau$  and let  $z_\beta$  be the single equilibrium macrostate of the corresponding Gibbs ensemble. Then there exists an  $\alpha > 0$  and an  $\varepsilon' > 0$  small enough such that for  $n$  large enough,*

$$\mathbb{E}_{\sigma, \tau}[d(X, Y)] \leq e^{-\alpha/n} d(\sigma, \tau)$$

whenever  $\|L_n(\sigma) - z_\beta\|_1 < \varepsilon'$ .

*Proof.* Let  $\varepsilon$  and  $\delta$  be as in Condition 6.12, and let  $\varepsilon' = \varepsilon^2 \delta / M$  with a constant  $M \gg 0$ .

**Case I.** Suppose  $L_n(\tau) = z$  and  $L_n(\sigma) = w$ , where  $\|z - z_\beta\|_1 \geq \varepsilon$  and  $\|w - z_\beta\|_1 < \varepsilon'$ .

Then there is an equilibrium configuration  $\sigma_\beta$  with  $L_n(\sigma_\beta) = z_\beta$  such that there is a *monotone path*

$$\pi' : \sigma_\beta = x'_0, x'_1, \dots, x'_r = \tau$$

connecting configurations  $\sigma_\beta$  and  $\tau$  on  $\Lambda^n$  such that  $\varepsilon \leq \|L_n(x'_i) - L_n(x'_{i-1})\|_1 < 2\varepsilon$ , and by Condition 6.12,

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x'_i) - L_n(x'_{i-1}), \nabla g_k^{H, \beta}(L_n(x'_{i-1})) \right\rangle \right|}{\|L_n(\sigma_\beta) - L_n(\tau)\|_1} \leq 1 - \delta/4$$

for  $n$  large enough. Note that the difference between the above inequality and Condition 6.12 is that here we take  $L_n(x'_i) \in \mathcal{P}_n$ .

Now, there exists a monotone path with from  $\sigma$  to  $\tau$

$$\pi : \sigma = x_0, x_1, \dots, x_r = \tau$$

such that

$$\|L_n(x_i) - L_n(x'_i)\|_1 \leq \varepsilon' \quad \text{for all } i = 0, 1, \dots, r.$$

The new monotone path  $\pi$  is constructed from  $\pi'$  by insuring that either

$$0 \leq \left\langle L_n(x_i) - L_n(x_{i-1}), e^k \right\rangle \leq \left\langle L_n(x'_i) - L_n(x'_{i-1}), e^k \right\rangle$$

or

$$\left\langle L_n(x'_i) - L_n(x'_{i-1}), e^k \right\rangle \leq \left\langle L_n(x_i) - L_n(x_{i-1}), e^k \right\rangle \leq 0$$

for  $i = 2, \dots, r$  and each coordinate  $k \in \{1, 2, \dots, q\}$ .

Then

$$\left| \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1} - \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x'_i) - L_n(x'_{i-1}), \nabla g_k^{H, \beta}(L_n(x'_{i-1})) \right\rangle \right|}{\|L_n(\sigma_\beta) - L_n(\tau)\|_1} \right| \leq C'' r \varepsilon' / \varepsilon$$

for a fixed constant  $C'' > 0$ . Noticing that  $r \varepsilon' / \varepsilon \leq \delta / M$  as  $r \leq 1/\varepsilon$ , and taking  $M$  large enough, we obtain

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1} \leq 1 - \delta/4.$$

Thus equation (38) will imply

$$\begin{aligned}
\mathbb{E}_{\sigma, \tau}[d(X, Y)] &\leq d(\sigma, \tau) \left[ 1 - \frac{1}{n} \left( 1 - \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right| + c\varepsilon^2}{\|L_n(\sigma) - L_n(\tau)\|_1} \right) \right] \\
&\leq d(\sigma, \tau) \left[ 1 - \frac{1}{n} (1 - (1 - \delta/4) - \delta/20) \right] \\
&= d(\sigma, \tau) \left[ 1 - \frac{1}{n} \delta/5 \right]
\end{aligned}$$

as  $\frac{c\varepsilon^2}{\|L_n(\sigma) - L_n(\tau)\|_1} \leq c\varepsilon \leq \delta/20$  for  $\varepsilon$  small enough.

**Case II.** Suppose  $L_n(\tau) = z$  and  $L_n(\sigma) = w$ , where  $\|z - z_\beta\|_1 < \varepsilon$  and  $\|w - z_\beta\|_1 < \varepsilon'$ .

Similarly to (37), equation (34) implies for  $n$  large enough,

$$\begin{aligned}
\mathbb{E}[d(X, Y)] &\leq d(\sigma, \tau) \cdot \left[ 1 - \frac{1}{n} \left( 1 - \frac{\|g^{H, \beta}(L_n(\sigma)) - g^{H, \beta}(L_n(\tau))\|_1}{\|L_n(\sigma) - L_n(\tau)\|_1} \right) \right] + O\left(\frac{1}{n^2}\right) \\
&\leq d(\sigma, \tau) \cdot \left[ 1 - \frac{\gamma}{n} \right] + O\left(\frac{1}{n^2}\right) \\
&\leq d(\sigma, \tau) \cdot \left[ 1 - \frac{\gamma}{2n} \right]
\end{aligned}$$

by Condition 6.13 (see also discussion following Condition 6.13).  $\square$

We now state and prove the main theorem of the paper that yields sufficient conditions for rapid mixing of the Glauber dynamics of the class of statistical mechanical models discussed.

**Theorem 6.15.** *Suppose  $H(z)$  and  $\beta > 0$  are such that Condition 6.12 and Condition 6.13 are satisfied. Then the mixing time of the Glauber dynamics satisfies*

$$t_{\text{mix}} = O(n \log n)$$

*Proof.* Let  $\varepsilon' > 0$  and  $\alpha > 0$  be as in Lemma 6.14. Let  $(X^t, Y^t)$  be a coupling of the Glauber dynamics such that  $X^0 \stackrel{\text{dist}}{=} P_{n, \beta}$ , the stationary distribution. Then, for sufficiently large  $n$ ,

$$\begin{aligned}
&\|P^t(Y^0, \cdot) - P_{n, \beta}\|_{\text{TV}} \\
&\leq P\{X^t \neq Y^t\} \\
&= P\{d(X^t, Y^t) \geq 1\} \\
&\leq \mathbb{E}[d(X^t, Y^t)] \\
&= \mathbb{E}[\mathbb{E}[d(X^t, Y^t) \mid X^{t-1}, Y^{t-1}]] \\
&\leq \mathbb{E}[\mathbb{E}[d(X^t, Y^t) \mid X^{t-1}, Y^{t-1}] \mid \|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'] \cdot P\{\|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'\} \\
&\quad + nP\{\|L_n(X^{t-1}) - z_\beta\|_1 \geq \varepsilon'\}.
\end{aligned}$$

By Lemma 6.14, we have

$$(39) \quad \begin{aligned} & \mathbb{E}[\mathbb{E}[d(X^t, Y^t) \mid X^{t-1}, Y^{t-1}] \mid \|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'] \\ & \leq e^{-\alpha/n} \mathbb{E}[d(X^{t-1}, Y^{t-1}) \mid \|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'] \end{aligned}$$

By iterating (39), it follows that

$$\begin{aligned} & \|P^t(Y^0, \cdot) - P_{n,\beta,K}\|_{\text{TV}} \\ & \leq e^{-\alpha/n} \mathbb{E}[d(X^{t-1}, Y^{t-1}) \mid \|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'] \cdot P\{\|L_n(X_{t-1}) - z_\beta\|_1 < \varepsilon'\} \\ & \quad + nP\{\|L_n(X^{t-1}) - z_\beta\|_1 \geq \varepsilon'\} \\ & \leq e^{-\alpha/n} \mathbb{E}[d(X^{t-1}, Y^{t-1})] + nP\{\|L_n(X^{t-1}) - z_\beta\|_1 \geq \varepsilon'\} \\ & \quad \vdots \\ & \leq e^{-\alpha t/n} \mathbb{E}[d(X^0, Y^0)] + n \sum_{s=0}^{t-1} P\{\|L_n(X^s) - z_\beta\|_1 \geq \varepsilon'\} \\ & = e^{-\alpha t/n} \mathbb{E}[d(X^0, Y^0)] + ntP_{n,\beta}\{\|L_n(X^0) - z_\beta\|_1 \geq \varepsilon'\} \\ & \leq ne^{-\alpha t/n} + ntP_{n,\beta}\{\|L_n(X^0) - z_\beta\|_1 \geq \varepsilon'\}. \end{aligned}$$

We recall the LDP limit (25) for  $\beta$  in the single phase region  $B$ ,

$$P_{n,\beta}\{L_n(X^0) \in dx\} \implies \delta_{z_\beta} \quad \text{as } n \rightarrow \infty.$$

Moreover, for any  $\gamma' > 1$  and  $n$  sufficiently large, by the LDP upper bound, we have

$$\begin{aligned} \|P^t(Y^0, \cdot) - P_{n,\beta}\|_{\text{TV}} & \leq ne^{-\alpha t/n} + ntP_{n,\beta}\{\|L_n(X^0) - z_\beta\|_1 \geq \varepsilon'\} \\ & < ne^{-\alpha t/n} + tne^{-\frac{n}{\gamma'} I_\beta(\varepsilon')}. \end{aligned}$$

For  $t = \frac{n}{\alpha}(\log n + \log(2/\varepsilon'))$ , the above right-hand side converges to  $\varepsilon'/2$  as  $n \rightarrow \infty$ .  $\square$

## 7. AGGREGATE PATH COUPLING APPLIED TO THE GENERALIZED POTTS MODEL

In this section, we illustrate the strength of our main result of section 6, Theorem 6.15, by applying it to the generalized Curie-Weiss-Potts model (GCWP), studied recently in [25]. The classical Curie-Weiss-Potts (CWP) model, which is the mean-field version of the well known Potts model of statistical mechanics [33] is a particular case of the GCWP model with  $r = 2$ . While the mixing times for the CWP model has been studied in [11, 27], these are the first results for the mixing times of the GCWP model.

Let  $q$  be a fixed integer and define  $\Lambda = \{e^1, e^2, \dots, e^q\}$ , where  $e^k$  are the  $q$  standard basis vectors of  $\mathbb{R}^q$ . A *configuration* of the model has the form  $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Lambda^n$ . We will consider a configuration on a graph with  $n$  vertices and let  $X_i(\omega) = \omega_i$  be the *spin* at vertex  $i$ . The random variables  $X_i$ 's for  $i = 1, 2, \dots, n$  are independent and identically distributed with common distribution  $\rho$ .



For the generalized Curie-Weiss-Potts model, for  $r \geq 2$ , the interaction representation function, defined in general in (3), has the form

$$H^r(z) = -\frac{1}{r} \sum_{j=1}^q z_j^r$$

and the generalized Curie-Weiss-Potts model is defined as the Gibbs measure

$$(40) \quad P_{n,\beta,r}(B) = \frac{1}{Z_n(\beta)} \int_B \exp \{ -\beta n H^r(L_n(\omega)) \} dP_n$$

where  $L_n(\omega)$  is the empirical measure defined in (21).

In [25], the authors proved that there exists a phase transition critical value  $\beta_c(q, r)$  such that in the parameter regime  $(q, r) \in \{2\} \times [2, 4]$ , the GCWP model undergoes a continuous, second-order, phase transition and for  $(q, r)$  in the complementary regime, the GCWP model undergoes a discontinuous, first-order, phase transition. This is stated in the following theorem.

**Theorem 7.1** (Generalized Ellis-Wang Theorem). *Assume that  $q \geq 2$  and  $r \geq 2$ . Then there exists a critical temperature  $\beta_c(q, r) > 0$  such that in the weak limit*

$$\lim_{n \rightarrow \infty} P_{n,\beta,r}(L_n \in \cdot) = \begin{cases} \delta_{1/q(1,\dots,1)} & \text{if } \beta < \beta_c(q, r) \\ \frac{1}{q} \sum_{i=1}^q \delta_{u(\beta,q,r)e^i + (1-u(\beta,q,r))/q(1,\dots,1)} & \text{if } \beta > \beta_c(q, r) \end{cases}$$

where  $u(\beta, q, r)$  is the largest solution to the so-called mean-field equation

$$u = \frac{1 - \exp(\Delta(u))}{1 + (q-1) \exp(\Delta(u))}$$

with  $\Delta(u) := -\frac{\beta}{q^{r-1}} [(1+(q-1)u)^{r-1} - (1-u)^{r-1}]$ . Moreover, for  $(q, r) \in \{2\} \times [2, 4]$ , the function  $\beta \mapsto u(\beta, q, r)$  is continuous whereas, in the complementary case, the function is discontinuous at  $\beta_c(q, r)$ .

For the GCWP model, the function  $g_\ell^{H,\beta}(z)$  defined in general in (30) has the form

$$g_k^{H,\beta}(z) = [\partial_k \Gamma](\beta \nabla H(z)) = [\partial_k \Gamma](\beta z) = \frac{e^{\beta z_k^{r-1}}}{e^{\beta z_1^{r-1}} + \dots + e^{\beta z_q^{r-1}}}.$$

For the remainder of this section, we will replace the notation  $H, \beta$  and refer to  $g^{H,\beta}(z) = (g_1^{H,\beta}(z), \dots, g_q^{H,\beta}(z))$  as simply  $g^r(z) = (g_1^r(z), \dots, g_q^r(z))$ . As we will prove next, the rapid mixing region for the GCWP model is defined by the following value.

$$(41) \quad \beta_s(q, r) := \sup \{ \beta \geq 0 : g_k^r(z) < z_k \text{ for all } z \in \mathcal{P} \text{ such that } z_k \in (1/q, 1] \}.$$

**Lemma 7.2.** *If  $\beta_c(q, r)$  is the critical value derived in [25] and defined in Theorem 7.1, then*

$$\beta_s(q, r) \leq \beta_c(q, r)$$

*Proof.* We will prove this lemma by contradiction. Suppose  $\beta_c(q, r) < \beta_s(q, r)$ . Then there exists  $\beta$  such that

$$\beta_c(q, r) < \beta < \beta_s(q, r).$$

Then, by Theorem 7.1, since  $\beta_c(q, r) < \beta$ , there exists  $u > 0$  satisfying the following inequality

$$(42) \quad u < \frac{1 - e^{\Delta(u)}}{1 + (q-1)e^{\Delta(u)}},$$

where  $\Delta(u) := -\frac{\beta}{q^{r-1}} [(1 + (q-1)u)^{r-1} - (1-u)^{r-1}]$ . Here, the above inequality (42) rewrites as

$$(43) \quad e^{\Delta(u)} = \exp \left\{ \beta \left[ \left( \frac{1-u}{q} \right)^{r-1} - \left( \frac{1+(q-1)u}{q} \right)^{r-1} \right] \right\} < \frac{1-u}{(q-1)u+1}.$$

Next, we substitute  $\lambda = (1-u)\frac{q-1}{q}$  into the above inequality (43), obtaining

$$(44) \quad \exp \left\{ \beta \left[ \left( \frac{\lambda}{q-1} \right)^{r-1} - (1-\lambda)^{r-1} \right] \right\} < \frac{\lambda}{(1-\lambda)(q-1)}.$$

Now, consider

$$z = \left( 1-\lambda, \frac{\lambda}{q-1}, \dots, \frac{\lambda}{q-1} \right).$$

Observe that  $z_1 = 1-\lambda = 1 - (1-u)\frac{q-1}{q} = \frac{1+u(q-1)}{q} > \frac{1}{q}$  as  $u > 0$ . Here, the inequality (44) can be consequently rewritten in terms of the above selected  $z$  as follows

$$z_1 = 1-\lambda < \frac{e^{\beta(1-\lambda)^{r-1}}}{e^{\beta(1-\lambda)^{r-1}} + (q-1)e^{\beta\left(\frac{\lambda}{q-1}\right)^{r-1}}} = g_1^r(z),$$

thus contradicting  $\beta < \beta_s(q, r)$ . Hence  $\beta_s(q, r) \leq \beta_c(q, r)$ .  $\square$

Combining Theorem 7.1 and Lemma 7.2 yields that for parameter values  $(q, r)$  in the continuous, second-order phase transition region  $\beta_s(q, r) = \beta_c(q, r)$ , whereas in the discontinuous, first-order, phase transition region,  $\beta_s(q, r)$  is strictly less than  $\beta_c(q, r)$ . This relationship between the equilibrium transition critical value and the mixing time transition critical value was also proved for the mean-field Blume-Capel model discussed in section 5. This appears to be a general distinguishing feature between models that exhibit the two distinct type of phase transition. We now prove rapid mixing for the generalized Curie-Weiss-Potts model for  $\beta < \beta_s(q, r)$  using the aggregate path coupling method derived in section 6.

We state the lemmas that we prove below, and the main result for the Glauber dynamics of the generalized Curie-Weiss-Potts model, a Corollary to Theorem 6.15. Let  $\beta_s(q)$  be the mixing time critical value for the GCWP model defined in (41).

**Lemma 7.3.** *Condition 6.11 and Condition 6.12 are satisfied for all  $\beta < \beta_s(q)$ .*

**Lemma 7.4.** *Condition 6.13 is satisfied for all  $\beta < \beta_s(q)$ .*

**Corollary 7.5.** *If  $\beta < \beta_s(q)$ , then*

$$t_{\text{mix}} = O(n \log n).$$

*Proof.* Condition 6.12 and Condition 6.13 required for Theorem 6.15 are satisfied by Lemma 7.3 and Lemma 7.4.  $\square$

*Proof of Lemma 7.4.* Denote  $z' = (z'_1, \dots, z'_q) = z - z_\beta$ . Then by Taylor's Theorem, we have

$$\begin{aligned}
 \limsup_{z \rightarrow z_\beta} \frac{\|g^r(z) - g^r(z_\beta)\|_1}{\|z - z_\beta\|_1} &= \limsup_{z \rightarrow z_\beta} \frac{\sum_{k=1}^q \left| \frac{e^{\beta z_k^{r-1}}}{\sum_{j=1}^q e^{\beta z_j^{r-1}}} - \frac{1}{q} \right|}{\sum_{k=1}^q \left| z_k - \frac{1}{q} \right|} \\
 &= \lim_{z' \rightarrow 0} \frac{\sum_{k=1}^q \left| \frac{\beta(r-1) \left(\frac{1}{q}\right)^{r-2} z'_k + O((z'_1)^2 + \dots + (z'_q)^2)}{q + O((z'_1)^2 + \dots + (z'_q)^2)} \right|}{\sum_{k=1}^q |z'_k|} \\
 (45) \qquad \qquad \qquad &= \frac{\beta(r-1) \left(\frac{1}{q}\right)^{r-2}}{q}.
 \end{aligned}$$

In [11], it was shown that for  $r = 2$ ,  $\beta < \beta_s(q) < \beta_c(q) < q$ . Therefore, the last expression above is less than 1 and we conclude that

$$\limsup_{z \rightarrow z_\beta} \frac{\|g^r(z) - g^r(z_\beta)\|_1}{\|z - z_\beta\|_1} < 1. \quad \square$$

*Proof of Lemma 7.3.* First, we prove that the family of straight lines connecting to the equilibrium point  $z_\beta = (1/q, \dots, 1/q)$  is a neo-geodesic family as it was defined following Condition 6.11. Specifically, for any  $z = (z_1, z_2, \dots, z_q) \in \mathcal{P}$  define the line path  $\rho$  connecting  $z$  to  $z_\beta$  by

$$(46) \qquad \qquad \qquad z(t) = \frac{1}{q}(1-t) + zt, \quad 0 \leq t \leq 1$$

Then, along this straight-line path  $\rho$ , the aggregate  $g$ -variation has the form

$$D_\rho^g(z, z_\beta) := \sum_{k=1}^q \int_\rho \left| \left\langle \nabla g_k^r(y), dy \right\rangle \right| = \sum_{k=1}^q \int_0^1 \left| \frac{d}{dt} [g_k^r(z(t))] \right| dt$$

Next, for all  $k = 1, 2, \dots, q$  and  $t \in [0, 1]$ , denote

$$z(t)_k = \frac{1}{q}(1-t) + z_k t$$

Then

$$(47) \qquad \qquad \qquad g_k^r(z(t)) = \frac{e^{\beta \left( \frac{1}{q}(1-t) + z_k t \right)^{r-1}}}{\sum_{j=1}^q e^{\beta \left( \frac{1}{q}(1-t) + z_j t \right)^{r-1}}}$$

and

$$(48) \quad \frac{d}{dt}[g_k^r(z(t))] = \beta(r-1)g_k^r(z(t)) \left[ \left( \frac{1}{q}(1-t) + z_k t \right)^{r-2} \left( z_k - \frac{1}{q} \right) - \langle z - z_\beta, g^r(z(t)) \rangle_\rho \right]$$

where  $\langle z - z_\beta, g^r(z(t)) \rangle_\rho$  is the weighted inner product

$$\langle z - z_\beta, g^r(z(t)) \rangle_\rho := \sum_{j=1}^q g_j^r(z(t)) \left( z_k - \frac{1}{q} \right) \left( \frac{1}{q}(1-t) + z_k t \right)^{r-2}$$

Now, observe that for  $z(t)$  as in (46) with  $z \neq z_\beta$ , the inner product  $\langle (z - z_\beta), g^r(z(t)) \rangle_\rho$  is monotonically increasing in  $t$  since

$$\frac{d}{dt} \langle z - z_\beta, g^r(z(t)) \rangle_\rho \geq \beta(r-1) \text{Var}_{g^r} \left( \left( z_k - \frac{1}{q} \right) \left( \frac{1}{q}(1-t) + z_j t \right)^{r-1} \right) > 0$$

where  $\text{Var}_{g^r}(\cdot)$  is the variance with respect to  $g^r$ .

So  $\langle z - z_\beta, g^r(z(t)) \rangle_\rho$  begins at  $\langle z - z_\beta, g^r(z(0)) \rangle_\rho = \langle z - z_\beta, z_\beta \rangle = 0$  and increases for all  $t \in (0, 1)$ .

The above monotonicity yields the following claim about the behavior of  $g_k^r(z(t))$  along the straight-line path  $\rho$ .

- (a) If  $z_k \leq 1/q$ , then  $g_k^r(z(t))$  is monotonically decreasing in  $t$ .
- (b) If  $z_k > 1/q$ , then  $g_k^r(z(t))$  has at most one critical point  $t_k^*$  on  $(0, 1)$ .

The above claim (a) follows immediately from (48) as  $\langle z - z_\beta, g^r(z(t)) \rangle_\rho > 0$  for  $t > 0$ . Claim (b) also follows from (48) as its right-hand side,  $z_k - 1/q > 0$  and  $\langle z - z_\beta, g^r(z(t)) \rangle_\rho$  is increasing. Thus there is at most one point  $t_k^*$  on  $(0, 1)$  such that  $\frac{d}{dt}[g_k^r(z(t))] = 0$ .

Next, define

$$A_z = \{k : z_k > 1/q\}$$

Then the aggregate  $g$ -variation can be split into

$$D_\rho^g(z, z_\beta) = \sum_{k \in A_z} \int_0^1 \left| \frac{d}{dt}[g_k^r(z(t))] \right| dt + \sum_{k \notin A_z} \int_0^1 \left| \frac{d}{dt}[g_k^r(z(t))] \right| dt$$

For  $k \notin A_z$ , claims (a) and (b) imply

$$\int_0^1 \left| \frac{d}{dt}[g_k^r(z(t))] \right| dt = - \int_0^1 \frac{d}{dt}[g_k^r(z(t))] dt = g_k^r(z(0)) - g_k^r(z(1)) = \frac{1}{q} - g_k^r(z)$$

For  $k \in A_z$ , let  $t_k = \max\{t_k^*, 1\}$ , where  $t_k^*$  is defined in (b). Then, we have

$$\int_0^1 \left| \frac{d}{dt}[g_k^r(z(t))] \right| dt = \int_0^{t_k^*} \frac{d}{dt}[g_k^r(z(t))] dt - \int_{t_k^*}^1 \frac{d}{dt}[g_k^r(z(t))] dt = 2g_k^r(z(t_k^*)) - g_k^r(z) - \frac{1}{q}$$

Combining the previous two displays, we get

$$\begin{aligned} D_\rho^g(z, z_\beta) &= \sum_{k \in A} \left( 2g_k^r(z(t_k^*)) - g_k^r(z) - \frac{1}{q} \right) + \sum_{k \notin A} \left( \frac{1}{q} - g_k^r(z) \right) \\ &= 2 \sum_{k \in A} \left( g_k^r(z(t_k^*)) - \frac{1}{q} \right) \end{aligned}$$

Since  $\beta < \beta_s$  and  $k \in A_z$ , we have

$$g_k^r(z(t_k^*)) < z(t_k^*)_k \leq z(1)_k = z_k$$

and we conclude that

$$D_\rho^g(z, z_\beta) < 2 \sum_{k \in A} \left( z_k - \frac{1}{q} \right) = \|z - z_\beta\|_1$$

Thus

$$\frac{d_g(z, z_\beta)}{\|z - z_\beta\|_1} \leq \frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} < 1 \quad \text{for all } z \neq z_\beta \text{ in } \mathcal{P}.$$

Next, since we are dealing with the straight line segments  $\rho$ ,

$$\limsup_{z \rightarrow z_\beta} \frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} = \limsup_{z \rightarrow z_\beta} \frac{\|g(z) - g(z_\beta)\|_1}{\|z - z_\beta\|_1} < 1$$

by (45), the Mean Value Theorem, and  $H(z) \in \mathcal{C}^3$ . This, in turn, guarantees the continuity required for Condition 6.11:

$$\limsup_{z \rightarrow z_\beta} \frac{d_g(z, z_\beta)}{\|z - z_\beta\|_1} \leq \limsup_{z \rightarrow z_\beta} \frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} < 1$$

Thus Condition 6.11 is proved for the CWP model. Moreover this proves that the family of straight line segments  $\rho$  is a neo-geodesic family (see definition following Condition 6.11). Indeed, there is  $\delta \in (0, 1)$  such that

$$\left\{ \rho : z(t) = \frac{1}{q}(1-t) + zt, \ z \in \mathcal{P} \right\} \quad \text{is a } \mathbf{NG}_\delta \text{ family of smooth curves,}$$

i.e.  $\forall z \neq z_\beta$  in  $\mathcal{P}$ , and corresponding  $\rho : z(t) = \frac{1}{q}(1-t) + zt$ ,

$$\frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} \leq 1 - \delta/2$$

Since the family of straight line segments  $\rho$  is a neo-geodesic family  $\mathbf{NG}_\delta$ , the integrals

$$D_\rho^g(x, z) := \sum_{k=1}^q \int_\rho \left| \left\langle \nabla g_k^r(y), dy \right\rangle \right|$$

can be uniformly approximated by the corresponding Riemann sums of small enough step size by the Mean Value Theorem as  $H(z) \in \mathcal{C}^3$  and therefore each  $g_k^r(z) \in \mathcal{C}^2$ . That is, there exists a constant  $C > 0$  that depends on the second partial derivatives of  $g^r(z) = (g_1^r(z), \dots, g_q^r(z))$ ,

such that for  $\varepsilon > 0$  small enough, the curve  $\rho = \frac{1}{q}(1-t) + zt$  in the family  $\mathbf{NG}_\delta$  that connects  $z_\beta$  to  $z$  satisfies

$$\left| \sum_{k=1}^q \sum_{i=1}^r \left| \left\langle z_i - z_{i-1}, \nabla g_k^r(z_{i-1}) \right\rangle \right| - D_\rho^g(z, z_\beta) \right| < Cr\varepsilon^2 \quad \forall z \in \mathcal{P} \text{ s.t. } \|z - z_\beta\|_1 \geq \varepsilon$$

for a sequence of points  $z_0 = z_\beta, z_1, \dots, z_r = z \in \mathcal{P}$  interpolating  $\rho$  such that

$$\varepsilon \leq \|z_i - z_{i-1}\|_1 < 2\varepsilon \quad \text{for } i = 1, 2, \dots, r.$$

Hence

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle z_i - z_{i-1}, \nabla g_k^r(z_{i-1}) \right\rangle \right|}{\|z - z_\beta\|_1} \leq 1 - \delta/2 + C\varepsilon \leq 1 - \delta/3$$

for  $\varepsilon \leq \delta/(6C)$ . This concludes the proof of Condition 6.12.  $\square$

## REFERENCES

1. N. Bhatnagar and D. Randall, Torpid mixing of simulated tempering on the Potts model, *Proceedings of the 15th annual ACM-SIAM symposium on Discrete algorithms*, 478–487 (2004).
2. M. Blume, Theory of the first-order magnetic phase change in  $\text{UO}_2$ , *Phys. Rev.* **141** (1966) 517–524.
3. M. Blume, V. J. Emery, and R. B. Griffiths, Ising model for the  $\lambda$  transition and phase separation in  $\text{He}^3\text{-He}^4$  mixtures, *Phys. Rev. A* **4** (1971) 1071–1077.
4. M. Bordewich and M.E. Dyer, Path coupling without contraction, *Journal of Discrete Algorithms*, Volume 5 Issue 2 (2007) 280–292.
5. R. Bubley and M.E. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science (FOCS)*, 223–231 (1997).
6. P. Brémaud, *Markov chains, Gibbs fields, Monte Carlo simulation, and queues*, Texts in Applied Mathematics, vol. 31, Springer-Verlag, New York, 1999.
7. H. W. Capel, On the possibility of first-order phase transitions in Ising systems of triplet ions with zero-field splitting, *Physica* **32** (1966) 966–988.
8. H. W. Capel, On the possibility of first-order phase transitions in Ising systems of triplet ions with zero-field splitting II, *Physica* **33** (1967) 295–331.
9. H. W. Capel, On the possibility of first-order phase transitions in Ising systems of triplet ions with zero-field splitting III, *Physica* **37** (1967) 423–441.
10. M. Costeniu, R. S. Ellis, and H. Touchette, Complete analysis of phase transitions and ensemble equivalence for the Curie-Weiss-Potts model. *Journal of Mathematical Physics*, Volume 46, 063301 (2005).
11. P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres and A. Sly, Glauber dynamics for the mean-field Potts model, *Journal of Statistical Physics*, Volume 149, Number 3, 432–477 (2012).
12. A. Dembo, and O. Zeitouni, *Large deviations techniques and applications*, Springer, New York, Second Edition (1998).
13. J. Ding, E. Lubetzky, and Y. Peres, The mixing time evolution of Glauber dynamics for the mean-field Ising model, *Comm. Math. Physics*, Volume 289, Number 2, 725–764 (2009).
14. J. Ding, E. Lubetzky, and Y. Peres, Censored Glauber dynamics for the mean-field Ising model, *Journal of Statistical Physics*, Volume 137, Number 1, 161–207 (2009).
15. M. Dyer, L.A. Goldberg, C. Greenhill, M. Jerrum, and M. Mitzenmacher, An extension of path coupling and its application to the Glauber dynamics for graph colorings, *SIAM Journal on Computing*, 30(6), 1962–1975 (2001).
16. T. Eisele and R. S. Ellis, Multiple phase transitions in the generalized Curie-Weiss model. *Journal of Statistical Physics*, Volume 52, Number 1/2, 161–207 (1988).
17. R. S. Ellis, *Entropy, Large Deviations and Statistical Mechanics*, Springer, New York, 1985. Reprinted in 2006 in *Classics in Mathematics*.
18. R. S. Ellis, K. Haven, and B. Turkington, Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles. *Journal of Statistical Physics*, Volume 101, Numbers 5/6, 999–1064 (2000).
19. M. Ebbens, H. Knöpfel, M. Löwe, and F. Vermet, Mixing times for the swapping algorithm on the Blume-Emery-Griffiths model, *Random Structures and Algorithms*, DOI:10.1002/rsa.20461 (2012).
20. M. Ebbens and M. Löwe, Rapid mixing of the swapping algorithm for the generalized Curie-Weiss model, submitted.
21. R. S. Ellis, J. Machta, and P. T. Otto. Asymptotic Behavior of the Magnetization Near Critical and Tricritical Points via Ginzburg-Landau Polynomials. *J. Stat. Phys.* **133** (2008) 101–129.
22. R. S. Ellis, J. Machta, and P. T. Otto. Asymptotic Behavior of the Finite-Size Magnetization as a Function of the Speed of Approach to Criticality. *Ann. Appl. Prob.* **20** (2010) 2118–2161.
23. R. S. Ellis, P. T. Otto, and H. Touchette, Analysis of phase transitions in the mean-field Blume-Emery-Griffiths model, *Ann. Appl. Prob.* **15** (2005) 2203–2254.
24. T.P. Hayes and E. Vigoda, Variable length path coupling, *Rand. Structures and Alg.*, Volume 31 Issue 3 (2007) 251–272.

25. B. Jahnel, C. Külske, E. Rudelli, and J. Wegener, Gibbsian and non-Gibbsian properties of the generalized mean-field fuzzy Potts-model, Accepted for publication in Markov Processes and Related Fields. Preprint arXiv:1312.5229v1
26. Y. Kovchegov, P. T. Otto, and M. Titus, Mixing times for the mean-field Blume-Capel model via aggregate path coupling, *Journal of Statistical Physics*, Volume 144, Numbers 5, 1009–1027 (2011).
27. Y. Kovchegov and P. T. Otto, Rapid Mixing of Glauber Dynamics of Gibbs Ensembles via Aggregate Path Coupling and Large Deviations Methods, preprint arXiv:1312.6728
28. D. A. Levin, M. Luczak, and Y. Peres, Glauber dynamics of the mean-field Ising model: cut-off, critical power law, and metastability, *Probability Theory and Related Fields*, Volume 146, Numbers 1, 223–265 (2010).
29. D. Levin, Y. Peres, E. Wilmer, Markov Chains and Mixing Times. American Mathematical Society, USA (2009)
30. T. Lindvall, Lectures on the Coupling Method, John Wiley & Sons, New York, 1992. Reprint: Dover paperback edition, 2002.
31. M. J. Luczak, Concentration of measure and mixing times of Markov chains. *Discrete Mathematics and Theor. Comp. Sci.* (Proceedings of the 5th Colloq. Mathem. Comp. Sci. ) 95–120 (2008).
32. J. Thompson, Optimal Couplings for Card Shuffling, *REU Program at Oregon State University Proceedings* (2010)
33. F. Y. Wu, The Potts model. *Rev. Mod. Phys.* 54:235-268 (1982).

DEPARTMENT OF MATHEMATICS, OREGON STATE UNIVERSITY, CORVALLIS, OR 97331  
*E-mail address:* kovchegy@math.oregonstate.edu

DEPARTMENT OF MATHEMATICS, WILLAMETTE UNIVERSITY, SALEM, OR 97302, PHONE: 1-503-370-6487  
*E-mail address:* potto@willamette.edu